```
     266 GlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGl 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     633 GGCCGCTTCATCCGCTGGCTCAACAAGGAGAAAGGCATCTTCAAAATTGA 584

     282 uAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProA 299
         ||||||||||||||||||||||||||||||:::||||||||||||||||
     583 GGACTCAGCACAGGTGGCCCGACTGTGGGGTGTGCGCAAGAACCGGCCAG 534

     299 laMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLys 315
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     533 CCATGAACTATGATAAACTAAGCCGCTCCATCCGCCAGTATTACAAGAAG 484

     316 GlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheVa 332
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     483 GGCATCATTCGTAAACCCGACATCTCTCAGCGCCTTGTCTACCAATTTGT 434

     332 lHisProIle 335
         |||||||:::
     433 GCATCCAGTC 424

seq_name: gb_est13:AA930371

seq_documentation_block:
LOCUS       AA930371      504 bp    mRNA            EST       23-APR-1998
DEFINITION  vo45a05.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone IMAGE:1052816 5' similar to SW:ETS4_DROME P29775 DNA-BINDING
            PROTEIN D-ETS-4 ;, mRNA sequence.
ACCESSION   AA930371
VERSION     AA930371.1  GI:3079964
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL , contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:584392
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1052816"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG], digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead. "
BASE COUNT      122 a    149 c    135 g     98 t
ORIGIN


alignment_scores:
           Quality:  548.00            Length:    104
             Ratio:  5.269               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 96.154

alignment_block:
 US-09-126-945B-2 x AA930371   ..

 Align seq 1/1  to: AA930371  from: 1  to: 504

     232 GluGluSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnPr 248
         |||||||:::||||||||||:::|||||||||||||||||||||||||||
       1 GAGGAGGGCTGGACGGATGGTGAGGTGGACTCGTCGTGCTCCGGGCAGCC 50

     248 oIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerT 265
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      51 CATTCACCTGTGGCAGTTCCTGAAAGAACTGCTGCTCAAGCCCCACAGCT 100

     265 yrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIle 281
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 ATGGCCGCTTCATCCGCTGGCTCAACAAGGAGAAAGGCATCTTCAAAATT 150

     282 GluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgPr 298
         ||||||||||||||||||||||||||||||:::|||||||||||||||||
     151 GAGGACTCAGCACAGGTGGCCCGACTGTGGGGTGTGCGCAAGAACCGGCC 200

     298 oAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysL 315
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 AGCCATGAACTATGATAAACTAAGCCGCTCCATCCGCCAGTATTACAAGA 250

     315 ysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 AGGGCATCATTCGTAAACCCGACATCTCTCAGCGCCTTGTCTACCAATTT 300

     332 ValHisProIle 335
         ||||||||||:::
     301 GTGCATCCAGTC 312

seq_name: gb_est89:BF535871

seq_documentation_block:
LOCUS       BF535871      637 bp    mRNA            EST       11-DEC-2000
DEFINITION  602051384F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4190637 5',
            mRNA sequence.
ACCESSION   BF535871
VERSION     BF535871.1  GI:11623239
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 637)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9518  row: q  column: 22
            High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1. .637
```

```
OM of: US-09-126-945P-2 to: EST:*   ext_format . pts

Date: Sep 18, 2001  6:43 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=tramev_p2n.model  DEV xlp
 -QY_[illegible]
 DB EST -QFMT fastap -SUFFIX rst -GAPOP=12.000 -GAPEXT 4.000
 MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP 4.500
 QGAPEXT 0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP 6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT 0.500 -DELOP=6.000
 -DELEXT 7.000 -START=1 -MATRIX blosum62 -TRANS=human40.rdi
 -LIST 150 -DOCALIGN 200 -THR_SCORE=pct -THR_MAX 100 -THR_MIN 0
 -ALIGN 100 -MODE LOCAL -OUTFMT pts -NORM=ext -MINLEN 0
 -MAXLEN=2000000000 -USER BS09126945_#CCN1_1_4005  NCPU 6
 -ICPU 3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2
Query length: 335
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 1364.200000
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ...
qb_est97:BG179568	+	1219.00	1992.39	6.5e-192	991	! BG179568 602428182F1 NIH_MGC_91
qb_est96:BG095247	+	997.00	1633.14	7.6e-92	591	! BG095247 B1110C12.5 NIA Mouse 1
qb_est78:BE757615	+	930.00	1523.63	9.5e-76	526	! BE757615 212104 MAGC 2B6V Pos 5
qb_est89:BF579116	+	849.50	1374.38	2.0e-67	2064	! BF579116 602096050F1 NCI_CGAP_S
qb_est96:BG072422	-	600.00	974.22	3.8e-45	784	! BG072422 B2110C12.3 NIA Mouse 1
qb_est13:AA930471	+	548.00	893.74	1.2e-40	504	! AA930471 ve45a05.r1 Barstead ma
qb_est89:BF535871	+	430.50	696.89	1.1e-29	637	! BF535871 602051384F1 NCI_CGAP_S
qb_est80:BF871340	+	425.00	683.01	6.3e-29	945	! BF871340 602134457F1 NIH_MGC_69
qb_est76:BE563238	+	409.00	656.74	1.8e-27	933	! BE563238 601335410F1 NIH_MGC_20
qb_qss5:CNS05AUV	+	399.50	639.47	1.6e-26	2021	! AL426858 Tetraodon nigroviridu
qb_est23:AI685592	+	392.00	639.42	1.7e-26	386	! AI685592 tc22c07.x1 NCI_CGAP_Pr
qb_est97:BG178499	+	363.00	582.03	2.7e-23	843	! BG178499 602028333F1 NIH_MGC_91
qb_est89:BF539428	+	340.50	543.40	3.8e-21	952	! BF539428 602094614F1 NCI_CGAP_S
qb_est24:AI745526	+	324.00	519.54	8.1e-20	721	! AI745526 wc81a10.x1 NCI_CGAP_Pr
qb_est10:AA662204	+	308.00	497.24	1.4e-18	514	! AA662204 ns66e07.r1 NCI_CGAP_Pr
qb_qss19:AQ977298	-	305.00	500.63	9.1e-19	250	! AQ977298 RPCI-23-421A14 TV RPCI
qb_qss5:CNS05KTR	+	304.50	483.40	8.3e-18	997	! AL341928 Tetraodon nigroviridis
qb_est99:BG328411	+	303.50	483.01	8.7e-18	899	! BG328411 602127746F1 NIH_MGC_17
qb_qss5:CNS04XXI	+	292.00	461.89	1.3e-16	1071	! AL422858 Tetraodon nigroviridi
qb_iss20:AI025433	-	280.00	453.64	3.8e-16	411	! AI025433 RPCI-23-345K9.TV RPCI
qb_est20:AI451780	+	278.50	447.84	7.9e-16	544	! AI451780 mk22107.y1 Soares_mous
qb_est54:AW988753	+	268.50	429.78	8.1e-15	618	! AW988753 CM3-CT0337.y1 Soares_mam
qb_htc:AK014602	+	264.50	462.15	2.8e-13	3480	! AK014602 Mus musculus 8 day em
qb_est90:BF580055	-	263.00	420.27	2.7e-14	640	! BF580055 ua20d04.x1 Gessler_Wil
qb_qss3:CNS027CY	+	257.00	405.45	1.6e-13	892	! AL145084 Tetraodon nigroviridis
qb_est97:BG172401	+	249.00	392.15	1.0e-12	966	! BG172401 602335688F1 NCI_CGAP_M
qb_est89:BF536831	+	242.50	386.64	2.0e-12	629	! BF536831 602048419F1 NCI_CGAP_S
qb_htc:AK018368	+	239.00	364.17	3.6e-11	2478	! AK018368 Mus musculus 16 days
qb_est83:BF133977	+	237.00	375.63	8.4e-12	737	! BF133977 601778421F1 NCI_CGAP_1
qb_est95:BF983301	+	234.00	373.06	1.2e-11	606	! BF983301 602305557F1 NIH_MGC_85
qb_est10:AA651760	+	233.50	373.88	1.0e-11	529	! AA651760 ns72a10.s1 NCI_CGAP_GC
qb_est13:AA913571	+	233.50	373.54	1.1e-11	544	! AA913571 ok31b10.s1 Soares_NFL_
qb_est99:BG327400	+	232.50	371.27	1.5e-11	656	! BG327400 602425871F1 NIH_MGC_24
qb_est81:BE951696	+	233.50	370.49	1.6e-11	705	! BE951696 UI-M-CC0-ayd-f-11-0-UI
qb_est53:AW944623	+	231.50	369.98	1.7e-11	556	! AW944623 EST186024 MAGE resequ
qb_est27:AI958988	+	231.50	369.85	1.8e-11	562	! AI958988 td06b10.y1 Zebrafish W
qb_est4:AA261341	+	230.00	367.52	2.4e-11	555	! AA261341 wa35d02.r1 Soares_mamm
qb_est100:BG394711	+	230.00	363.26	4.1e-11	788	! BG394711 602121214F1 NIH_MGC_93
qb_est29:AL560620	+	228.50	359.19	6.9e-11	898	! AL560620 AL560620 LLI_NFL010_B2
qb_est88:BF450651	+	228.00	362.45	4.5e-11	642	! BF450651 qz88g07.y1 NCI_CGAP_Ma
qb_est80:BF888101	+	227.50	362.67	4.4e-11	589	! BF888101 [illegible]F1 NIH_MGC_73
qb_est1:AA229986	+	226.50	361.54	5.5e-11	441	! AA229986 nc50d12.r1 NCI_CGAP
qb_est87:BF379908	+	225.50	367.54	2.4e-11	301	! BF379908 CM4-UT0044-260900 5
qb_est84:BF180927	+	225.50	353.41	1.4e-10	961	! BF180927 601807167F1 NCI_CGA
qb_est94:BF953438	+	224.00	358.60	7.4e-11	512	! BF953438 BC3_NN0195-121100 0
qb_est31:AU129451	+	224.00	355.25	1.1e-10	674	! AU129451 AU129451 NI2BL2 Homo
qb_est29:AL556433	+	224.00	352.46	1.6e-10	848	! AL556433 AL556433 LLI_NFL006
qb_est29:AL555439	+	224.00	350.41	2.1e-10	1003	! AL555439 AL555439 LLI_NFL 00
qb_est89:BF584991	+	223.50	351.45	1.9e-10	861	! BF584991 602098856F1 NCI_CGA
qb_est96:BG069473	-	223.00	351.73	1.8e-10	786	! BG069473 B3075B09.3 NIA Mouse
qb_est88:BF376484	-	222.00	352.86	1.6e-10	622	! BF376484 nc50d12.r1 NCI_CGA
qb_est54:AW985653	+	221.00	354.45	1.3e-10	479	! AW985653 qi84b02.y1 Soares_m
qb_est87:BF443190	+	221.00	354.38	1.3e-10	482	! BF443190 2B0593 MAGC 2B6V Su
qb_est111:W98560	+	221.00	352.66	1.6e-10	555	! W98560 wa14e04.r1 Soares_mou
qb_est96:BG076722	+	220.50	351.84	1.8e-10	555	! BG076722 B3005A08.5 NIA Mous
qb_est100:BG394169	+	220.50	348.91	4.9e-10	1065	! BG394169 602456455F1 NIH_MG
qb_est78:BE764557	-	219.50	352.16	1.4e-10	424	! BF764557 PM4_NI0073_2305 0
qb_est29:AL560651	+	218.00	350.94	7.2e-10	968	! AL560651 AL560651 LLI_NFL010
qb_est45:AW280428	+	217.50	346.73	8.4e-10	562	! AW280428 tj40d01.y1 zebrafis
qb_est53:AW911816	+	217.00	345.63	8.9e-10	575	! AW911816 qi39c06.y1 Soares_m
qb_est27:AI991109	+	216.00	349.48	8.6e-10	842	! AI991109 wu38c03.x1 Soares_b
qb_est73:BE314959	+	215.50	342.31	6.0e-10	616	! BE314959 601144e4e4F1 NIH_MGC
qb_est46:AW111287	-	215.50	341.85	6.4e-10	640	! AW111287 tj11d10.y1 NIH_MGC_
qb_est95:BF976822	+	215.50	338.83	9.4e-10	820	! BF976822 602146661F1 NIH_MGC
qb_est77:BE627411	-	215.00	340.74	7.2e-10	655	! BE627411 qt94c09.x1 Soares_m
qb_est38:AV405665	+	214.00	336.31	1.3e-09	823	! AV405665 AV405665 Bambyx mori
qb_est77:BE616829	+	214.00	341.00	1.7e-09	995	! BE616829 601280947F1 NIH_MGC
qb_est80:BE884422	+	213.50	344.49	1.6e-09	863	! BE884422 601529174F1 NIH_MGC
qb_est98:BG294409	+	213.00	334.75	1.6e-09	817	! BG294409 602395158F1 NIH_MGC
qb_est80:BE859344	+	212.50	341.43	6.7e-10	441	! BE859344 UI-M-AQ0-aaf-c-04-0
qb_est28:AL120951	+	212.50	341.29	6.8e-10	446	! AL120951 DKFZp762E043_r1 762
qb_est76:BE553236	+	212.50	337.24	1.1e-09	622	! BE553236 qr43c06.y1 NCI_CGAP
qb_est71:BE200487	-	212.00	340.07	1.3e-09	640	! BE200487 qq63c08.x1 Soares_m
qb_est47:AW464228	-	211.50	338.42	9.9e-10	492	! AW464228 PF230015A10H1 Soare
qb_est26:AI877251	+	211.00	333.02	2.0e-09	738	! AI877251 te51b11.y1 Zebrafis
qb_est89:BF571541	+	210.50	330.10	2.9e-09	853	! BF571541 602078149F1 NIH_MGC
qb_est84:BF144998	+	210.00	331.12	2.5e-09	733	! BF144998 601790295F1 NCI_CGA
qb_est10:C46450	+	209.50	338.64	9.6e-10	369	! C46450 C46450 Yuji Kohara un
qb_est94:BF896156	-	209.50	338.38	9.9e-10	377	! BF896156 CM2-MT0187-231100-5
qb_est77:BE626845	-	209.50	337.57	1.1e-09	403	! BE626845 qt88c01.x1 Soares_p
qb_est26:AI876285	+	209.50	337.01	1.2e-09	422	! AI876285 qz73d07.y1 Soares_f
qb_est106:N29134	+	209.50	334.95	1.5e-09	509	! N29134 yx47c05.r1 Soares_mel
qb_est8:BE901513	-	209.50	334.56	1.6e-09	510	! BE901513 qp48d11.y1 NCI_CGAP
qb_est31:AU140178	+	209.50	328.50	3.6e-09	863	! AU140178 AC14178 FinalFA Hom
qb_est52:AW853070	+	209.00	333.85	1.8e-09	511	! AW853070 BCL-CT0249-030300 0
qb_est52:AW853081	+	209.00	332.74	2.0e-09	560	! AW853081 BCL-CT0249-030300 0
qb_est52:AW853019	+	209.00	331.52	2.4e-09	619	! AW853019 BCL-CT0249-170200 0
qb_est98:BG304069	+	208.00	328.62	3.5e-09	686	! BG304069 [illegible]xl Soares_f
qb_est26:AI891569	+	207.50	332.06	2.2e-09	484	! AI891569 qt31q06.x1 Soares_m
qb_est84:BF148041	+	207.50	329.75	3.0e-09	584	! BF148041 qw93c06.x1 Soares_m
qb_est84:BF148053	+	207.50	329.71	3.0e-09	586	! BF148053 qw93d07.x1 Soares_m
qb_est83:BF117201	+	207.50	329.42	3.1e-09	600	! BF117201 qy96c11.x1 NCI_CGAP
qb_est100:BG388964	+	207.50	322.17	7.9e-09	1089	! BG388964 602414721F1 NIH_MG
qb_est29:AL527893	+	206.50	325.06	5.5e-09	750	! AL527893 AL527893 LLI_NFL003
qb_est79:BF778347	+	206.00	325.70	5.1e-09	665	! BF778347 602134746F1 NIH_MGC
qb_est86:BF520959	+	205.50	326.08	1.8e-09	682	! BF520959 qz57e08.y1 NCI_CGAP
qb_est86:BF844411	+	205.50	320.28	1.0e-08	970	! BF844411 602014802F1 NCI_CGA
qb_est69:BF526348	+	205.00	328.01	4.2e-09	424	! BF526348 qb34e04.y1 Stratoki
qb_est87:BF440199	+	205.00	324.80	5.2e-09	625	! BF440199 7F4310 Y Soares_NFL
qb_est87:BF440177	+	204.50	325.96	4.9e-09	581	! BF440177 BS2e02[illegible]220 Jewish
qb_est46:AW429841	+	203.50	327.88	3.8e-09	396	! AW429841 EST330 MAGE liver Sus
qb_est92:BF769441	+	203.50	325.09	5.5e-09	498	! BF769441 BC3-110010 111[illegible]
qb_est83:BF080551	+	203.50	324.42	6.0e-09	526	! BF080551 231500 MAGC 2B6V Su
qb_est73:BE310161	+	203.50	322.96	7.2e-09	592	! BE310161 601092574F1 NCI_CGA
qb_est101:BG542246	+	203.50	319.99	1.1e-08	752	! BG542246 602561356F1 NIH_MGC
qb_est47:AW462268	+	202.50	326.95	4.4e-09	372	! AW462268 PF230012A2dH12 Soar
qb_est47:AW461923	+	202.50	326.01	4.9e-09	403	! AW461923 PF230007A10H1 Soare
qb_est77:BE686024	+	202.50	324.38	6.0e-09	461	! BE686024 150401 MAGC 4B6V Bo
qb_est90:BF503021	+	202.50	324.96	5.6e-09	477	! BF503021 268425 MAGC 5B6V Bo
qb_est75:BE485600	+	202.50	324.86	6.4e-09	481	! BE485600 172685 MAGC 5B6V Bo
qb_est32:AI573169	+	202.50	322.74	7.5e-09	525	! AI573169 [illegible]xl NCI_CGAP
qb_est88:BE904106	+	202.50	315.65	1.8e-08	944	! BE904106 601494[illegible]F2 NIH_MGC
qb_est97:BG173029	+	202.00	311.74	2.3e-08	952	! BG173029 602345510F1 NCI_CGA
qb_est19:AA543917	+	201.50	329.24	1.5e-08	614	! AA543917 nk43e12.r1 Soares_ga

```
qb_est75:BE508079     +  200.50  820.41  1.0e-08   487  ! BE508079 de26e09.y1 NIHBD XGC
qb_est27:AI498666     +  200.50  815.56  1.9e-08   725  ! AI498666 AI498666 dkf2426 Call
qb_est91:BF697665     +  200.50  810.80  3.4e-08  1072  ! BF697665 602131202F1 NIH_MGC_9
qb_est149:AW579379    +  200.00  819.83  1.1e-08   477  ! AW579379 PM0_DT0065-120100-001
qb_est101:BG470225    +  200.00  816.05  1.7e-08   651  ! BG470225 6825837F9F1 NIH_MGC_19
qb_est41:AU129884     +  200.00  815.64  1.8e-08   673  ! AU129884 AU129884 NT2RP2 Homo s
qb_est28:AL525441     +  200.00  814.14  2.5e-08   827  ! AL525441 AL525441 LTI_NFL003_NH
qb_est29:AL555767     +  200.00  812.63  2.7e-08   862  ! AL555767 AL555767 LTI_NFL006_PI
qb_est15:AI037444     +  199.50  825.02  5.5e-09   291  ! AI037444 qb52f12.x1 Soares_mamm
qb_est103:C42136      +  199.50  822.43  7.7e-09   360  ! C42136 C42136 Yuji Kohara unpub
qb_est76:BE563956     +  199.50  814.57  2.1e-08   687  ! BE563956 601348167F1 NIH_MGC_54
qb_est25:AI793162     +  199.00  816.46  1.7e-08   554  ! AI793162 oh51e09.y5 NCI_CGAP_Co
qb_est25:AI791558     +  199.00  816.44  1.7e-08   555  ! AI791558 oh86f09.y5 NCI_CGAP_Co
qb_est50:AW657616     +  199.00  816.21  1.7e-08   561  ! AW657616 110841 MARC IPIG Sus s
qb_est89:BF584050     +  199.00  813.83  2.3e-08   682  ! BF584050 602096296F1 NCI_CGAP_t
qb_est40:AV733666     +  199.00  813.24  2.5e-08   716  ! AV733666 AV733666 cDA Homo sapi
qb_est13:AA919938     +  198.50  817.59  1.4e-08   468  ! AA919938 vy18h08.r1 Stratagene
qb_est84:BF181270     +  198.50  816.69  1.6e-08   504  ! BF181270 601805754F1 NCI_CGAP_M
qb_est21:AI558312     +  198.50  816.10  1.7e-08   529  ! AI558312 IL78102.x1 Zebrafish W
qb_est12:AA815785     +  198.50  815.51  1.9e-08   555  ! AA815785 vv13e03.r1 Stratagene
qb_est49:AW601848     +  198.50  814.44  2.2e-08   611  ! AW601848 QV4-BT0407-260100 087-
qb_est84:BF143562     +  198.50  811.42  3.2e-08   777  ! BF143562 601790916F1 NCI_CGAP_L
qb_est84:BF140129     +  198.50  809.04  4.3e-08   945  ! BF140129 601786079F1 NCI_CGAP_L
qb_est44:AV192171     +  197.50  819.13  1.2e-08   360  ! AV192171 AV192171 Yuji Kohara u
qb_est78:BE715674     -  197.50  818.54  1.3e-08   378  ! BE715674 CM0-BT0747-190600-433
qb_est17:AI175900     -  197.50  816.20  1.7e-08   458  ! AI175900 EST219472 Normalized r
qb_est90:BF603437     +  197.50  815.68  1.8e-08   478  ! BF603437 268948 MARC HBOV Bos t
qb_est93:BF982778     -  197.50  815.55  1.9e-08   483  ! BF982778 QV3-ET0176-061200-515
qb_est52:AW841181     -  197.50  814.68  2.1e-08   519  ! AW841181 RC1-CN0012-120200-012
qb_est71:BE191776     -  197.50  814.28  2.5e-08   582  ! BE191776 db82e01.x1 Wellcome CR
qb_est29:AL587041     +  197.50  812.05  2.9e-08   644  ! AL587041 AL587041 Stratagene CH
qb_est102:BG547609    +  197.50  811.81  3.0e-08   657  ! BG547609 602565944F1 NIH_MGC_77
qb_est96:BG093894     -  197.50  811.18  3.3e-08   692  ! BG093894 nu91f02.x1 Soares_mous
qb_qss4:CNS04ORL      +  197.50  807.87  5.0e-08   908  ! AL500378 Tetraodon nigroviridis
qb_est29:AL571262     -  197.50  807.53  5.2e-08   934  ! AL571262 AL571262 LTI_NFL006_PI
qb_est100:BG420378    +  197.50  806.99  5.6e-08   976  ! BG420378 6024524F7F1 NIH_MGC_14
```

```
seq name: qb_est97:BG179568

seq documentation block:
LOCUS        BG179568       991 bp    mRNA             EST       06 FEB 2001
DEFINITION   602428182F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429413 5',
             mRNA sequence.
ACCESSION    BG179568
VERSION      BG179568.1  GI:12686271
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 991)
  AUTHORS    NIH MGC http://mgc.nci.nih.gov/
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10181  row: d  column: 22
             High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1..991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429413"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT       190 a    351 c    305 g    145 t
ORIGIN
```

```
alignment_scores:
            Quality: 1219.00              Length:    278
              Ratio:  4.686                 Gaps:      7
 Percent Similarity:  93.525   Percent Identity: 88.849

alignment_block:
 US-09-126-945B-2.x-BG179568   ..

 Align seq 1/1  to: BG179568  from: 1  to: 991

      34 GlyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaTh 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       3 GGGGCAGTGGGTCTCGAGAGACGGGACTGGAGTCCCAGTCCACCCGCCAC 52

      50 rProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuT 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      53 GCCCGAGGCAGGGCCTGTCCGCCTTCTACTCTCCTACTTTGACATGCTGT 102

      67 yrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArg 83
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     103 ACCCTGAGGACAGCAGCTGGGCAGCCAAGGCCCCTGGGGCCAGCAGTCGG 152

      84 GluGluProProGluGluProGluGlnCysProValIleAspSerGlnAl 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     153 GAGGAGCCACCTGAGGAGCCTGAGCAGTGCCCGGTCATTGACAGCCAAGC 202

     100 aProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluH 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     203 CCCAGCGGGCAGCCTGGACTTGGTGCCCGGCGGGCTGACCTTGGAGGAGC 252

     117 isSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLysAsp 133
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     253 ACTCGCTGGAGCAGGTGCAGTCCATGGTGGTGGGCGAAGTGCTCAAGGAC 302

     134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTr 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     303 ATCGAGACGGCCTGCAAGCTGCTCAACATCACCGCAGATCCCATGGACTG 352

     150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGluTyrA 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     353 GAGCCCCAGCAATGTGCAGAAGTGGCTCCTGTGGACAGAGCACCAATACC 402

     167 rqLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeu 183
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     403 GGCTGCCCCCATGGGCAAGGCTTCCAGGAGCTGGCGGGCAAGGAGCTG 452

     184 CysAlaMetSerGluGlnPheArgGlnArgSerProLeuGlyAs 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     453 TGCGCCATGTCGGAGGAGCAGTTCCGCCAGCGCTCGCCCCTGGGTGGGGA 502

     200 pValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysG 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     503 TGTGCTGCACGCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAAAG 552

     217 luArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGln 233
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     553 AGCGGACTTCACCTGGGGCGATTCACTACTGTGCCTCGACCAGTGAGGAG 602

     234 SerThrIleAspSerGluValAspSerSerCysSerGlyGlnProIleHi 250
         ||||||||||||||::||||||||||||||||||||||||||||||||||||
     603 AGCTGGACCGACAGCCGAGTGGACTCATCATGCTCCGGGCAGCCA...TC 649

     250 sLeuTrpGlnPheLeuLysGlnLeuLeuLeuLysProHisSerTyrGly 266
         ||||||||||||||||||::||||||||||||   ||||||||||||
```

```
      650 CACTGTGGCAGTTCCTCAAGCAGTTGCTACTCAAGCC.CACAGCTATGGC 698

      267 ArgPheIleArgTrpLeuAsn.LysGluLysGlyIlePheLysIleGluA 283
          ||||||||||          :::    |||:::|||:::|||   ||||||
      699 CGCTTCATAAGTGGC...TCACCACGAGAACGGCCTCTTCCACATTGAGG 745

      283 spSerAlaGlnValAlaArgLeuTrp.GlyIleArgLysAsnArgProAl 299
          |||||   :::      |||   |||    |||   ::::::||||||
      746 ACTCACGCCACCGCGCCCCC...TCGGTGCATCCGACCTACCCCCCC.. 790

      299 aMetAsnTyrAspLysLeuSerArgSer 308
          |||::::::::::::|||:::   |||
      791 .ATGAGATTCGAACAGCTGACCGGCTCC 817

seq_name: qb_est96:BG085247

seq_documentation_block:
LOCUS       BG085247      591 bp    mRNA            EST       26-JAN-2001
DEFINITION  H3110C12-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3110C12 5', mRNA sequence.
ACCESSION   BG085247
VERSION     BG085247.1  GI:12567811
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Karpul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jamakit,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3110C12-3
            Contact: George J. Karpul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15K.html for details.
            Plate: H3110  row: C  column: 12
            Seq primer: -21M13 Reverse
            High quality sequence stop: 591
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..591
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="H3110C12"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /sex="Clones arrayed from a variety of cDNA libraries"
                     /dev_stage="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (including preimplantation stage
                     embryos from  unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad)  and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with oligo(dT
                     )-Not primers.  References include: (1) Genome-wide
                     expression profiling of mid-gestation  placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray,  2000,
                     Proc. Natl. Acad. Sci. U.S.A. 97: 9127-9132; (2)
                     Large-scale cDNA  analysis reveals phased gene expression
                     patterns during preimplantation  mouse development, 2000,
                     Development, 127: 1737-1749; (3) Genome wide  mapping of
                     unselected transcripts from extraembryonic tissue of
                     7.5-day  mouse embryos reveals enrichment in the t-complex
                     and under-representation  on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT       135 a    168 c    175 g    114 t
ORIGIN


alignment_scores:
            Quality:  997.00              Length:     195
              Ratio:  5.220                 Gaps:       0
 Percent Similarity:  97.949   Percent Identity: 92.821

alignment_block:
 US-09-126-945B-2 x BG085247   ..

 Align seq 1/1  to: BG085247  from: 1  to: 591

      141 LeuAsnIleThrAlaAspProMetAspIrpSerProSerAsnValGluLy 187
          ||||||||||||||||||||   ||||||||||||:::|||||||||||
        1 CTGAACATCACAGCAGACCCTGGGGACTGGAGCCCTGGTAACGTGCAGAA 50

      187 sTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysA 174
          |||||||||||||||||||||||||||||||||||||||||   |||||||
       51 GTGGCTTTTATGGACAGAACACACATACCGGCTGCCTCCAGCAGGCAAGG 100

      174 laPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGln 190
          ||||||||||:|||:::|||||||||||||||||||:::||||||||||||
      101 CCTTCCAGGAGCTGGGCGGTAAGGAGCTGTGCGCTTTGTCCGAGGAACAG 150

      191 PheArgGluArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAs 207
          ||||||||||:::|||||||||||||||||||||||||||||||||||||
      151 TTCCGTCAGCGTGCACCCTTGGGTGGGCATGTACTGCATGCCCAGCTGGA 200

      207 pIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAla1 224
          |||||||||||||||||||||||||||||||||||||||||||||||:::
      201 CATCTGGAAGTCAGCGGCCTGGATGAAGGAGAGGACCTCGCCTGGGACCC 250

      224 leHisTyrLysAlaSerThrSerGlnGlnSerTrpThrAspSerGlnVal 240
          ::|||||||||||||||||||||||||||:::|||||||||:::||||||
      251 TTCACTACTGCGCCTCCACCAGCGAGCAGCGCTGGACGGATGCTGAGGTG 300

      241 AspSerSerLysSerSerIleIleProTyrGluLeuTrpGlnPheLeuLys 257
          ||||||||||||||||||||||||||||.|||   |||   ||||||||
      301 GACTCGTCGAAGTCTCCGGGCAGCCCATTCACCTGGGCAGGTTCTGAAAGA 350

      257 uLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnL 274
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      351 ACTGCTGCTCAAGCCCCACAGCTATGGCCGCTTCATCCGCTGGCTCAACA 400

      274 ysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeu 290
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      401 AGGAGAAAGGCATCTTCAAAATTGAGGACTCAGCACAGGTGGCCCGAGTG 450

      291 TrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerAr 307
          ||||||:::|||||||||||||||||||||||||||||||||||||||||
      451 TGGGGTGTGCGCAAGAACCGGCCAGCCATGAACTATGATAAACTAAGCCG 500

      307 qSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleS 324
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      501 CTCCATCCGCCAGTATTACAAGAAGGGCATCATTCGTAAACCCGACATCT 550

      324 erGlnArgLeuValTyrGlnPheValHisProIle 345
          ||||||||||||||||||||||||||||||||:::
      551 CTCAGCGCCTTGTCTACCAATTTGTGCATCCAGTC 585

seq_name: qb_est78:BE757615

seq_documentation_block:
LOCUS       BE757615      549 bp    mRNA            EST       15-SEP-2000
DEFINITION  212104 MARC 2B6V Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE757615
VERSION     BE757615.1  GI:10171607
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing  Bases called and clip_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the  minscore 18
            and  minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 64  row: o  column: 13
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..539
                     /organism "Bos taurus"
                     /db_xref "taxon:9913"
                     /clone_lib "MARC ZBOV"
                     /tissue_type "pooled"
                     /lab_host "DH10B"
                     /note "Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT       98 a    168 c    193 g     80 t
ORIGIN


alignment_scores:
            Quality:  930.00             Length:    178
              Ratio:  5.225                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 95.506

alignment_block:
 US-09-126-945B-2 x BE757615   ..

 Align seq 1/1  to: BE757615  from: 1  to: 539

      85 GluProProGluGluProGluGlnCysProValIleAspSerGlnAlaPr 101
         ||||||||||:::||||||||||||||||||||||||||||||||||||||
          5 GAGCCGCCTGACGAGCCCGAGCAGTGCCCGGTCATCGACAGCCAAGCCCC 54

     101 nAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHis: 118
         |:::||||||||||||||||:::|||||||||||||||||||||||||||
          55 AGGGGGCAGCCTGGACCTGGCCCCGGGCGGGCTGACCCTGGAGGAGCACT 104

     118 erLeuGluGlnValGluSerMetValValGlyGluValLeuLysAspIle 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
         105 CGCTGGAGCAGGTGCAGTCCATGGTGGTGGGCGAGGTGCTCAAGGACATC 154

     135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
         ||||||||||||||||||||||||||||||||||||||||:::||||||||
         155 GAGACGGCCTGCAAGCTGCTCAACATCACGGCAGACCCTGTGGACTGGAG 204

     151 rProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 168
         ||||:::||||||||||||||||||||||||||||||||||||||||||||
         205 CCCTGGCAACGTGCAGAAGTGGCTGCTGTGGACGGAGCACCAGTACCGGC 254

     168 euProProMetGlyLysAlaPheGluGluLeuAlaGlyLysGluLeuCys 184
         |||||||||:::||||||||||||||||||||||:::||||||||||||||
         255 TGCCCCCCGTGGGCAAGGCCTTCCAGGAGCTGGGGGGCAAGGAGCTGTGT 304

     185 AlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspVa 201
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
         305 GCCATGTCCGAGGAGCAGTTCCGCCAGCGCTCGCCCCTGGGCGGGGACGT 454

     201 lLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluA 218
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
         455 GCTGCACGCCCATCTGGACATCTGGAAATCAGCTGCCTGGATGAAAGAGA 404

     218 rgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSer 234
         ||||||||||||||||||||||||:::||||||||||||||||||||||||
         405 GGACGTCCCCCGGGGGCAATTCACTTCTGCGCCTGGACCAGCCAGGAGAGC 454

     235 TrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHisLe 251
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
         455 TGGACGGACAGCGAGGTGGACTCGTCCTGCTCCGGGCAGCCCATCCACCT 504

     251 uTrpGlnPheLeuLysGluAsnLeuLysPro 262
         |||||||||||||||||||||||||||||||
         505 CTGGCAGTTTCTCAAGGAGCTGCTGCTGAAGCCG 538

seq_name: qb_est89:BE579116

seq_documentation_block:
LOCUS       BE579116      2064 bp    mRNA             EST       12 DEC 2000
DEFINITION  602996050F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4215901 5',
            mRNA sequence.
ACCESSION   BE579116
VERSION     BE579116.1  GI:11652828
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2064)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9792  row: d  column: 14
            High quality sequence stop: 645.
FEATURES             Location/Qualifiers
     source          1..2064
                     /organism "Mus musculus"
                     /strain "FVB/N"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:4215901"
                     /clone_lib "NCI_CGAP_Co24"
                     /lab_host "DH10B (T1 phage-resistant)"
                     /note "Organ: colon; Vector: pCMV SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      624 a    621 c    620 g    199 t
ORIGIN


alignment_scores:
            Quality:  849.50             Length:    245
              Ratio:  3.988                Gaps:      5
 Percent Similarity:  86.939   Percent Identity: 76.327

alignment_block:
 US-09-126-945B-2 x BE579116   ..

 Align seq 1/1  to: BE579116  from: 1  to: 2064
```



```
seq_name: gb_est96:BG072422

seq_documentation_block:
LOCUS       BG072422      784 bp    mRNA            EST       25-JAN-2001
DEFINITION  H3110C12-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3110C12 3', mRNA sequence.
ACCESSION   BG072422
VERSION     BG072422.1  GI:12554991
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 784)
  AUTHORS   Karpul,G.J., Pudenkia,D.E., Qian,Y., Lim,M.K., Tarabar,S.A.,
            Tsuka,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3110C12-5
            Contact: George J. Karqul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for more details.
            Plate: H3110  row: C  column: 12
            Seq primer:  21M13 Forward
            High quality sequence stop: 784
            POLYA Yes.
FEATURES             Location/Qualifiers
     source          1..784
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="H3110C12"
                     /clone_lib "NIA Mouse 15K cDNA Clone Set"
                     /note "Clones arrayed from a variety of cDNA libraries"
                     /dev_stage "Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo cDNA libraries (including preimplantation stage
                     embryos from unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos, extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad) and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with oligo(dT
                     ) Not primers. References include: (1) Genome-wide
                     expression profiling of mid-gestation placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray, 2000,
                     Proc. Natl. Acad. Sci. U.S.A. 97: 9127-9132; (2)
                     Large-scale cDNA analysis reveals phased gene-expression
                     patterns during preimplantation mouse development, 2000,
                     Development, 127: 1737-1749; (3) Genome-wide mapping of
                     unselected transcripts from extraembryonic tissue of
                     7.5-day mouse embryos reveals enrichment in the t-complex
                     and under-representation on the X chromosome, 1999, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT      156 a    215 c    223 g    190 t
ORIGIN


alignment_scores:
            Quality:  699.00              Length:    720
              Ratio:  5.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 94.167

alignment_block:
 US-09-126-945B-2 x BG072422/rev  ..

 Align seq 1/1  to reverse of: BG072422  from: 1  to: 784

     216 LysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGl 242
         ||||||||||||||||||||:::||||||||||||||||||||||||||
     782 AAGGAcAGGAcCTCGCcTGGcACCcTccAcTACTGCGCcTCTcCAGCGA 733

     232 aGluSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnPro1 249
         ||||:::|||||||||:::|||||||||||||||||||||||||||||
     732 GGAGGCTGGACGGATGCTGAGTGGACTCGTCGTGCTCCGGGCAGCCCA 683

     249 leHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyr 265
         ||||||||||   ||||||||||||||||||||||||||||||||||||
     682 TTCACCTGTGGAG.TTCCTGAAAGAACTGCTGCTCAAGCCCCACAGCTAT 634
```

```
      266 GlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGl 282
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      633 GGCCGCTTCATCCGCTGGCTCAACAAGGAGAAAGGCATCTTCAAAATTGA 584

      282 uAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProA 299
          ||||||||||||||||||||||||||||||||:::||||||||||||||||
      583 GGACTCAGCACAGGTGGCCCGACTGTGGGGTGTGCGCAAGAACCGGCCAG 534

      299 laMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLys 315
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      533 CCATGAACTATGATAAACTAAGCCGCTCCATCCGCCAGTATTACAAGAAG 484

      316 GlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheVa 332
          ||||||||||||||||||||||||||||||||||||||||||||||||
      483 GGCATCATTCGTAAACCCGACATCTCTCAGCGCCTTGTCTACCAATTTGT 434

      332 lHisProIle 335
          ||||||||:::
      433 GCATCCAGTC 424

seq_name: qb_est13:AA930471

seq_documentation_block:
LOCUS        AA930471      504 bp    mRNA            EST       23-APR-1998
DEFINITION   vo45a05.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
             clone IMAGE:1052816 5' similar to SW:ETS4_DROME P29775 DNA-BINDING
             PROTEIN D-ETS-4 ;, mRNA sequence.
ACCESSION    AA930471
VERSION      AA930471.1  GI:3079964
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 504)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberq,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
  TITLE      The WashU HHMI Mouse EST Project
  JOURNAL    Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
             WashU HHMI Mouse EST Project
             Washington University School of MedicineP
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.qov) for further information.
             MGI:584492
             Seq primer:  28ml3 rev2 ET from Amersham
             High quality sequence stop: 496.
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Mus musculus"
                     /strain "FVB/N"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:1052816"
                     /clone_lib "Barstead mouse irradiated colon MPLRB7"
                     /dev_stage "8 weeks"
                     /lab_host "DH10B"
                     /note "Vector: pT7T3D Pac (Pharmacia) with a modified
                     polylinker; Site 1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Cys.  1st strand cDNA was primed
                     with  a Not I   oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGCAGCGGCCGCCTTTTTTTTTTTTTTTTTTTTT
                     T 4'];  double stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG],  digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead.  "
BASE COUNT       122 a    149 c    135 q     98 t
ORIGIN


alignment_scores:
            Quality:  548.00              Length:    104
              Ratio:  5.269                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 96.154

alignment_block:
 US-09-126-945B-2 x AA930471   ..

 Align seq 1/1  to: AA930471  from: 1  to: 504

      232 GluGluSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnPr 248
          ||||||:::||||||||:::||||||||||||||||||||||||||||||
        1 GAGGAGGGCTGGACGGATGGTGAGGTGGACTCGTCGTGCTCCCGGGCAGCC 50

      248 olleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerT 265
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 CATTCACCTGTGGCAGTTCCTGAAAGAACTGCTGCTCAAGCCCCACAGCT 100

      265 yrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIle 281
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      101 ATGGCCGCTTCATCCGCTGGCTCAACAAGGAGAAAGGCATCTTCAAAATT 150

      282 GluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgPr 298
          ||||||||||||||||||||||||||||||||:::||||||||||||||
      151 GAGGACTCAGCACAGGTGGCCCGACTGTGGGGTGTGCGCAAGAACCGGCC 200

      298 oAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysL 315
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      201 AGCCATGAACTATGATAAACTAAGCCGCTCCATCCGCCAGTATTACAAGA 250

      315 ysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      251 AGGGCATCATTCGTAAACCCGACATCTCTCAGCGCCTTGTCTACCAATTT 300

      332 ValHisProIle 335
          ||||||||||:::
      301 GTGCATCCAGTC 312

seq_name: qb_est89:BF535871

seq_documentation_block:
LOCUS        BF535871      637 bp    mRNA            EST       11-DEC-2000
DEFINITION   602051384F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4190647 5',
             mRNA sequence.
ACCESSION    BF535871
VERSION      BF535871.1  GI:11623239
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 637)
  AUTHORS    N.H.M.C. [...] ;, [...] ;D. [...],
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberq, Ph.D.
             Email: cqapbs-r@mail.nih.qov
             Tissue Procurement: Jeffrey E. Green, M.D.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.qov
             Plate: LLAM9518  row: q  column: 22
             High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1. .637
```

```
                     /organism "Mus musculus"
                     /strain "129/SvJ"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:4190637"
                     /clone_lib="NCI_CGAP_Sg2"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note "Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo
                     dT. Average insert size 1.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT       120 a    219 c    179 g    114 t      5 others
ORIGIN


alignment_scores:
            Quality:  440.50              Length:    141
              Ratio:    3.776               Gaps:      6
 Percent Similarity:   80.851   Percent Identity: 68.794

alignment_block:
 US-09-126-945B-2 x BF535871   ..

 Align seq 1/1  to: BF535871  from: 1  to: 637

       1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
         ||||||||||||||||||||||||||||||:|||||||||||:::   |||| |
     248 ATGGGCAGTGCCAGCCCAGCCTGAGCAACGTGTCCCCCGCTTGCCTGCT 297

      17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 64
         ||||   |||||||:::::   |||||||||   |||||||||||||:::|
     298 ACTGTTCCCAGATGTGGCACCACGAACAGGGACCGCAGAAGCCAGCATCAG 347

      64 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
         ||||||:|||   ||||:::::::|||||||||||:|||||||||||||
     348 GAGCAATGGGCCCTGAGAACCAGGAATGGAGTCCTAGTCCACCCCCAGC 397

      51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
         |||||||||||||||||||||||||||||||||||||||:|||   ||
     398 CCTGAGCAGGGCCTGTCTGCTTTCTACCTCTCTTACTTTAACATG...TA 444

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
         ||||:::|||||||||||||:::||||||:::|||   |||   :::   |
     445 TCCCGACGATAGCAGCTGGGTCGCCAAAGTCCCCGAGGCCCGTGCCGGGG 494

      84 luGluProProGluGlu.ProGlu.....GlnCys.Pro...ValIleA 97
         :|:::   |||||||||| ||||||         ||| |||      |||
     495 AGGACCACCCGGAGGAGGCCCGAAAGNCAAGNNTGTCCCCNGTTTCATTG 544

      97 spSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThr 113
         ||||||||||||                        :::|||   |||
     545 ACAGCCAGGCC.........................TCTGGGAGCACG 567

     114 LeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
         |||:::|||||||||||||||||||||||||||||||||||||||||||||
     568 TTGGATGAGCACTCGCTAGAGCAGGTGCAATCGATGGTNGTGGGCGAGGT 617

     130 ILeuLysAspIleGluThr 136
         |||||||||||||||||
     618 CCTGAAAGATATTGAGACG 636

seq_name: gb_est80:BE871340

seq_documentation_block:
LOCUS       BE871340      945 bp    mRNA            EST       22-OCT-2000
DEFINITION  601449457F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3853334 5',
            mRNA sequence.
ACCESSION   BE871340
VERSION     BE871340.1  GI:10320116
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9577  row: a  column: 15
            High quality sequence stop: 656.
FEATURES             Location/Qualifiers
     source          1..945
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="IMAGE:3853334"
                     /clone_lib "NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host "DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT       170 a    363 c    400 g    112 t
ORIGIN


alignment_scores:
            Quality:  425.00              Length:    152
              Ratio:    3.484               Gaps:      4
 Percent Similarity:   80.263   Percent Identity: 69.737

alignment_block:
 US-09-126-945B-2 x BE871340   ..

 Align seq 1/1  to: BE871340  from: 1  to: 945

       1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     465 ATGGGCAGCGCCAGCCCGGGTCTGAGCAGCGTATCCCCCAGCCACCTCCT 414

      17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 64
         ||||||||||||||||   ||||||||||||||||||||||||   ||||
     415 GCTGCCCCCCGACACGGG.TGGCGGACAGGCTTGGAGAAGGCGGA.GCGG 462

      64 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
         ||||||:|||||||||||||||||||||||||||||||||||||||||||
     463 GGGCAGCGGGTCTCGAGAGACGGGACTGGAGTCCCAGTCCACCCGCCACG 512

      51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
         |||:::::   :::      ||||||||||||||||      ||||||
     513 CCGAGCAGGCCTGTC...CGCTTCTACCTCTCCTACTCGACCATGCTGTC 559

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
         ||||||||||||||||||||||||:::      |||   :::|||||||
     560 ACCTGAGGACAGCAGCTGGGCAGCAAGGCCCCTGGGGCCAGCCAGTCGGG 609

      84 luGluProProGluGluProGluGlnCysProValIleAspSerGlnAla 100
         |||||||||||:|||||||||||||||   ||||||:::   :::   :::
     610 AGGAGCCACCTGAGGAGCCTGAGCAGGGCCCGGTCATG...ACAGCAAGC 656

     101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHi 117
         ||:|||:::::|||   ||||||||||||||   ||||||||||||||
     657 CCAGCTGCCGCAGCTGCACTTGGTGCCCGGCGGCTG.ACCTTGGAGGAGCA 705

     117 sSerLeuGluGlnVal.GlnSerMetValValGlyGluValLeuLysAsp 133
         ||||||||||||||||   ||||||||||||||||||||
     706 CTCGCTGGAGCAGCTGGCCGTCCATGGTGGTGGGCGAGGTGCTCAAGAC 755
```

```
      144 IleGlyThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIr 150
          ||||||                 :::   :::   :::|||   :::||
      756 TCCGAGACGCCT...........GAAGCGTCACAATCACCGCACTCCTG 793

      150 pSer 151
          |:::
      794 GACT 797

seq name: qb_est76:BE563288

seq documentation block:
LOCUS       BE563288       933 bp    mRNA             EST       15 AUG 2000
DEFINITION  601345410F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3689402 5',
            mRNA sequence.
ACCESSION   BE563288
VERSION     BE563288.1  GI:9807100
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 933)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM384  row: c  column: 03
            High quality sequence stop: 678.
FEATURES             Location/Qualifiers
     source          1..933
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3689402"
                     /clone_lib="NIH_MGC_39"
                     /tissue_type "adenocarcinoma"
                     /lab_host "DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
                     Site 2: EcoRI; cDNA made by oligo dT priming;
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies)."
BASE COUNT       148 a    416 c    419 g    150 t
ORIGIN


alignment scores:
            Quality:  409.00             Length:    117
              Ratio:  3.971                Gaps:      2
 Percent Similarity:  88.034   Percent Identity: 84.615

alignment block:
 US-09-126-945B-2 x BE563288  ..

 Align seq 1/1  to: BE563288  from: 1  to: 933

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      395 ATGGGCAGCGCCAGCCCGGGTCTGAGCAGCGTATCCCCCAGCCACCTCCT 444

       17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAla. 33
          ||||||||||||||||||||||||||||||||||||||||||||||||
      445 GCTGCCCCCCCGACACGGTGTCGCGGACAGGCTTGGAGAAGCCGGCAGCGC 494

       34 GlyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaTh 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      495 GGGGCAGTGGGTCTGGACAGACGGGACTGGAGTCCCAGTCCACCCGCCAC 544

       50 rProGluGlnGly.LeuSerAlaPheTyrLeuSerTyr.PheAspMetLe 66
          |||||||||||||  :::       ||||||||||||||| ||||||||||
      545 GCCCGAGCAGGGCTGTC...CGCTTCTACCTCTCCTACTTTGACATGCT 591

       66 ulyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerA 83
          ||||||||||||||||||   ::::::|||||||||||   |||||||
      592 GTACCCTGAGGACAGCAGCCTGGGCAGCAAGGCCCCTGGGGCAGCAGTC 640

       83 rqGluGluProProGluGluPro.GluGlnCysProValIleAsp..... 97
          |||||||||||||||||||||| ||||||||||||||||||||
      641 GGGAGGAGCCACCTGAGGAGCCTTGAGCAGTGCCCGGTCATTGACTGACA 690

       98 .SerGlnAlaProAlaGlySerLeuAspLeuValProGlyGly 111
          :::|||     ||||||          ||||||||||||||
      691 AGCCCAGCGCCAGGCTGCA.........TTGGTGCCGGGCGGG 724

seq name: qb_gss5:CNS05AUV

seq documentation block:
LOCUS       CNS05AUV      1020 bp    DNA              GSS       26 MAY 2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            027021 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL329008
VERSION     AL329008.1  GI:8222630
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1020)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1020)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1020)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12 APR 2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..1020
                     /organism "tetraodon nigroviridis"
                     /db_xref "taxon:99883"
                     /clone "027021"
                     /clone_lib "A"
                     /note "Genoscope sequence ID : CDAA027AH11A1 end : 1"
BASE COUNT       298 a    215 c    218 g    273 t     16 others
ORIGIN


alignment scores:
            Quality:  399.50             Length:    252
              Ratio:  3.995                Gaps:      3
```

```
 Percent Similarity:  89.683   Percent Identity:  86.111

alignment_block:
 US-09-126-945B-2 x CNS05AUV   ..

 Align seq 1/1  to: CNS05AUV  from: 1  to: 1020

     232 GluGluSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnPr 248
         ||| ..:|||       |||||:::|||||||||||||||||||||||
      10 GAGNNGTGNTGG......TCCGAGCCCGATTCTTCCTGCTCCGGGCAGCC 58

     248 gIleHisLeuTrpGlnPheLeuLysGluLeuLeuLysProHisSerI 265
         |||||||||||||||||||||||:::|||||||||||||||||||||:::|
      54 GATCCATCTGTGGCAGTTCCTCCGAGAACTCTTACTCAAACCCCACAACC 103

     265 yrGlyArgPheIleArgTrpLeuAsnLysGlu.Lys..... 276
         ||||||||   ||||||||||||||||||:: |||
     104 ACGGGCGCTGCATCCGCTGGCTGAACAAGGACAAAGGTATCTTCACCTGT 153

     276 ..................................................  276

     154 TTCCTGCATGTTTACATTTTAAAGATAGATGTTCTAACAGTGTGCCTGGA 203

     276 ..................................................  276

     204 GCGAGGACGTGATCTGCACCACCAATGCAGATTAAACAATGAAAGAAAAC 253

     276 ..................................................  276

     254 ATCTCAAATGCATGTTTCTTTTATTAAATTCCCATATTTATTAAGACACT 303

     276 ..................................................  276

     304 GTATACATTTGAAAAAAAAAAAAATTAGTTGTGGTAAATATGACTGTTGT 353

     276 ..................................................  276

     354 TTGATATTATTCTTAGTATAAATATTAATTTAAACGAGATTCTTTCAAAC 403

     276 ..................................................  276

     404 TGGAATATCATAGAAACTCCAAAACAGATGTGTTTTGTCATTTGTTTGT 453

     276 ..................................................  276

     454 ATTTTTTTAAGAGTACGTCTGAAATTCTAGTTTAAATAAAATGAAAACCT 503

     276 ..................................................  276

     504 CAGCCAAAGACATGAAAGACATCATCCTGGAACTTGTTTCACTGTTAGCA 553

     277 .........................GlyIlePheLysIleGluAspSe 284
                                 ||||||||||||||||||||||
     554 AAAACCCACCGTCTTATTGTTGCTTCAGGTATTTTTAAAATTGAGGACTC 603

     284 rAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetA 301
         ||||:::|||||||||||||||||||:::|||||||||||||||||||||
     604 GGCTCACGTGGCCAGGCTTTGGGGCCTCCGGAAGAATCGACCTGCCAIGA 653

     301 snIyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIle 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     654 ACTACGACAAGCTGAGCCGCTCGATACGTCAGTACTACAAGAAGGGGATC 703

     318 IleArgLysProAspIleSerGlnArgLeuValTyrGlnPheValHisPr 334
         ||||||||||||||||:::|||||||||||||||||||||||||||||||
     704 ATTCCCAAGCCCGATGTGTCCCAGACGTGGTCTACCAGTTCGTGCACCC 753

     334 oIle 335
         ||||
     754 AGTG 757

seq_name: gb_est23:AI685592
```

```
seq_documentation_block:
LOCUS       AI685592       386 bp    mRNA            EST       15-DEC-1999
DEFINITION  tx20c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
            similar to contains PTR5.t3 MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI685592
VERSION     AI685592.1  GI:4896886
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   [illegible]
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1143  Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 380.
FEATURES             Location/Qualifiers
     source          1..386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251596"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT        72 a    140 c    111 g     64 t
ORIGIN
```

```
alignment_scores:
           Quality:  392.00              Length:     75
             Ratio:  5.227                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2 x AI685592   ..

 Align seq 1/1  to: AI685592  from: 1  to: 386

       1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     161 ATGGGCAGCGCCAGCCCGGGTCTGAGCAGCGTATCCCCCAGCCACCTCCT 210

      17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     211 GCTGCCCCCCCGACACGGTGTCGCGGACAGGCTTGGAGAAGGCGGCAGCGG 260

      34 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     261 GCGCACTGGGTCTCCAGACACGGGACTGGAGTCCCAGTCCACCCGCCACG 310
```

```
      51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     411 CCCGAGCAGGGCCTGTCCGCCTTCTACCTCTCCTACTTTGACATGCTGTA 360

      67 rProGluAspSerSerTrpAlaAla 75
         |||||||||||||||||||||||||
     361 CCCTGAGGACAGCAGCTGGGCAGCC 385

seq_name: qb_est97:BG178499

seq_documentation_block:
LOCUS        BG178499       843 bp    mRNA             EST       06 FEB 2001
DEFINITION  602428333F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4429712 5',
            mRNA sequence.
ACCESSION    BG178499
VERSION      BG178499.1  GI:12685214
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 843)
  AUTHORS    NIH_MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10182  row: a  column: 09
             High quality sequence stop: 686.
FEATURES             Location/Qualifiers
     source          1..843
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4429712"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage resistant)"
                     /note="Stratagene pCMV-SPORT6, Cla I - Not I,
                     Site_2: Sall; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched by
                     full length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      210 a    260 c    243 g    130 t
ORIGIN


alignment_scores:
            Quality:  864.00            Length:     69
              Ratio:  5.261               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2 x BG178499   ..

 Align_seq 1/1  to: BG178499  from: 1  to: 843

     267 ArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAs 283
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CGCTTCATTAGGTGGCTCAACAAGGAGAAGGGCATCTTCAAAATTGAGGA 51

     284 pSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaM 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 CTCAGCCCAGGTGGCCCGGCTGTGGGGCATCCGCAAGAACCGTCCCGCCA 101

     600 etAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGly 616
```

```
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 TGAACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAAGGGC 151

     317 IleIleArgLysProAspIleSerGluArgLeuValTyrGlnPheValHi 333
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 ATCATCCGGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCA 201

     333 sProIle 335
         |||||||
     202 CCCCATC 208

seq_name: qb_est89:BF539428

seq_documentation_block:
LOCUS        BF539428       952 bp    mRNA             EST       11 DEC 2000
DEFINITION  602049614F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4189178 5',
            mRNA sequence.
ACCESSION    BF539428
VERSION      BF539428.1  GI:11626809
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 952)
  AUTHORS    NIH_MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Jeffrey E. Green, M.D.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM9514  row: k  column: 04
             High quality sequence stop: 636.
FEATURES             Location/Qualifiers
     source          1..952
                     /organism="Mus musculus"
                     /tissue="VENN"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4189178"
                     /clone_lib="NCI_CGAP_SG2"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="organ, salivary gland, Vector: pCMV SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo
                     dT. Average insert size 1.8 kb. Constructed by Life
                     Technologies. Note: this is a NCI CGAP Library."
BASE COUNT      183 a    312 c    298 g    158 t      1 others
ORIGIN


alignment_scores:
            Quality:  340.50            Length:    100
              Ratio:  3.869               Gaps:      1
 Percent Similarity:  88.000   Percent Identity: 71.000

alignment_block:
 US-09-126-945B-2 x BF539428   ..

 Align_seq 1/1  to: BF539428  from: 1  to: 952

       1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
         ||||||||||||||||||||||||:::|||||||||||:::   |||||
     390 ATGGGCAGTGCCAGCCCAGGCCTGAGCAACGTGTCCCCCGGTTGCCTGCT 439

      17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 34
         ||||   |||||:::::   |||||||||   |||||||||||||:::|
     440 ACTGTTCCCAGATGTGGCACCACGAACAGGGACGGACAAGGCAGCATCAG 489
```

```
      84 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
         ||||||::|||   |||::::::::::||||||||||||||||||||||
     490 GAGCAATGGGCCCTGACAAGCACGAATGGAGTCCTAGTCCACCCGCCAGC 539

      51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuLy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     540 CCTGAGCAGGGCCTGTCTGCTTTCTACCTCTCTTACTTTAACATG...TA 586

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
         ||||::|||||||||||||::||||||::|||   |||   :::   |
     587 PCCCGACGATAGCAGCTGGGTCGCCAAAGTCCCCGAGGCCCGTGCGGC.G 635

      84 lnGluProProGluGluProGluGlnCysProValIleAspSerGlnAla 100
         ||::::   ||||||||||||||::::::      ||||||||||||||||
     636 ACGACCAACCGGAGGAGCCGAGCAGGGTGGCCGTCATTGACAGCCAGGCT 685

seq_name: qb_est24:AI745526

seq_documentation_block:
LOCUS       AI745526      721 bp    mRNA            EST       17-DEC-1999
DEFINITION  we34a10.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2417050 3'
            similar to SW:ETS4_DROME P29775 DNA-BINDING PROTEIN D ETS-4 ;, mRNA
            sequence.
ACCESSION   AI745526
VERSION     AI745526.1  GI:5113814
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 826   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 448.
FEATURES             Location/Qualifiers
     source          1. .721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone "IMAGE:2417050"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT713D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clone IDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      122 a    197 c    241 g    159 t      2 others
ORIGIN


alignment_scores:
            Quality:  424.00            Length:     83
              Ratio:  4.050               Gaps:      0
 Percent Similarity:  96.386   Percent Identity: 92.771

alignment_block:
 US-09-126-945B-2 x AI745526/rev  ..

 Align seq 1/1  to reverse of: AI745526  from: 1  to: 721

     253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl 269
         ||||||||||||||:::||||||||||:|||   ||||||   |||||||
     721 CAGTTCCTCAAGAGTTTGCTACTCAGCCCNNC.AGCTATGNCCGCTTCAT 673

     269 cArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaC 286
         ||||||||||:::|||||||||||||||   ||||||||||||||||||
     672 AAGGTGGCTNCAAAGGAGAAGGGCATCTTC.AAAATTGAGGACTCAGCCC 624

     286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     623 AGGTGGCCCGCCTGTGGGGCATCCGCAAGAACCGTCCCGCAATGAACTAC 574

     303 AspLysIleSerArgSerIleArgGluTyrTyrLysLysGlyIleIleAr 319
         ||||||||||||||||||||||||||||   ||||||||||||||||||
     573 GACAAGCTGAGCCGCTCCATCCGCGAGTA.TACAAGAAGGGCATCATCCG 525

     319 qLysProAspIleSerGlnArgLeuValTyrGlnPheValHisProIle 335
         |||||||||||||:|||||||||||||||||||||||||||||||||||
     524 NAAGCCAGACATCTCNCAGCGCCTCGTCTACCAGTTCGTGCACCCCATC 476

seq_name: qb_est10:AA662204

seq_documentation_block:
LOCUS       AA662204      514 bp    mRNA            EST       03-DEC-1997
DEFINITION  nsf66c07.r1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1188588 5'
            similar to SW:ETS4_DROME P29775 DNA-BINDING PROTEIN D ETS-4 ;, mRNA
            sequence.
ACCESSION   AA662204
VERSION     AA662204.1  GI:2616295
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 744   Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .514
                     /organism "Homo sapiens"
                     /db_xref: "taxon:9606"
                     /clone "IMAGE:1188588"
                     /clone lib "NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type "normal prostate"
                     /lab_host "DH10B"
                     /note "Organ: prostate; Vector: pT713D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I  and Eco RI sites of the
```

```
                    Modified pIN13 vector.  Library is normalized, and was
                    constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       108 a    174 c    151 g     82 t
ORIGIN


alignment scores:
            Quality:  408.00              Length:      60
              Ratio:  5.133                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 98.333

alignment block:
 US-09-126-945B-2 x AA662204  ..

 Align seq 1/1  to: AA662204  from: 1  to: 514

      226 LysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGl 292
          :::|||||||||||||||||||||||||||||||||||||||||||||||||
        2 GAGGGCATCTTCAAAATTGAGGACTCAGCCCAGGTGGCCCGGCTGTGGGG 51

      293 yIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerI 309
          |||||||||||||||||||||||||||||||||||||||||||||||||
       52 CATCCGCAAGAACCGTCCCGCCATGAACTATGACAAGCTGAGCCGCTCCA 101

      309 leArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGln 325
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      102 TCCGCCAGTATTACAAGAAGGGCATCATCCGCAAGCCAGACATCTCCCAG 151

      326 ArgLeuValTyrGlnPheValHisProIle 335
          ||||||||||||||||||||||||||||||
      152 CGCCTCGTCTACCAGTTCGTGCACCCCATC 181

seq_name: qb_qss19:AQ977298

seq_documentation_block:
LOCUS        AQ977298       259 bp    DNA              GSS       29-JAN-2000
DEFINITION   RPCI-23 321A14.TV RPCI-23 Mus musculus genomic clone RPCI-23-321A14
             , DNA sequence.
ACCESSION    AQ977298
VERSION      AQ977298.1  GI:6809599
KEYWORDS     GSS.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 259)
  AUTHORS    Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shetsman,S., Akinsot
             ,B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
             and Fraser,C.M.
  TITLE      Mouse BAC End Sequences from Library RPCI-23
  JOURNAL    Unpublished (1999)
COMMENT      Other GSSs: RPCI-23 321A14.TJ
             Contact: Shaying Zhao
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0200
             Fax: 301 838 0208
             Email: szhao@tigr.org
             Clones are derived from the mouse BAC library RPCI-23. For BAC
             library availability, please contact Pieter de Jong
             (pieter@dejong.med.buffalo.edu). Clones may be purchased from
             BACPAC Resources (http://bacpac.med.buffalo.edu/orderinfo.htm)
             or from Roswell Genetics (info@roswellgen.com). BAC end pages:
             http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
             Plate: 321  row: A  column: 14
             Seq primer: T7
             Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..259
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23:321A14"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT        52 a     60 c     85 g     62 t
ORIGIN


alignment scores:
            Quality:  405.00              Length:      59
              Ratio:  5.169                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 96.610

alignment block:
 US-09-126-945B-2 x AQ977298/rev  ..

 Align seq 1/1  to reverse of: AQ977298  from: 1  to: 259

      277 GlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIl 293
          ||||||||||||||||||||||||||||||||||||||||||||||||::
      249 GGCATCTTCAAAATTGAGGACTCAGCACAGGTGGCCCGACTGTGGGGTGT 200

      293 cArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleA 310
          :|||||||||||||||||||||||||||||||||||||||||||||||||
      199 GCGCAAGAACCGGCCAGCCATGAACTATGATAAACTAAGCCGCTCCATCC 150

      310 rgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArg 326
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      149 GCCAGTATTACAAGAAGGGCATCATTCGTAAACCCGACATCTCTCAGCGC 100

      327 LeuValTyrGlnPheValHisProIle 335
          |||||||||||||||||||||||:::
       99 CTTGTCTACCAATTTGTGCATCCAGTC 73

seq_name: qb_qss5:CNS05K1K

seq_documentation_block:
LOCUS        CNS05K1K       997 bp    DNA              GSS       26-MAY-2000
DEFINITION   Tetraodon nigroviridis genome survey sequence 17 end of clone
             042E03 of library A from Tetraodon nigroviridis, genomic survey
             sequence.
ACCESSION    AL341928
VERSION      AL341928.1  GI:8235686
KEYWORDS     GSS; genome survey sequence.
SOURCE       Tetraodon nigroviridis.
  ORGANISM   Tetraodon nigroviridis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
             Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
             Tetraodontidae; Tetraodon.
REFERENCE    1  (bases 1 to 997)
  AUTHORS    Roest Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
             Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
             Weissenbach,J.
  TITLE      Characterization and repeat analysis of the compact genome of the
             freshwater pufferfish Tetraodon nigroviridis
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 997)
  AUTHORS    Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
             Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
             Saurin,W. and Weissenbach,J.
  TITLE      Human gene number estimate provided by genome wide analysis using
             Tetraodon nigroviridis DNA sequence
  JOURNAL    Unpublished
REFERENCE    3  (bases 1 to 997)
  AUTHORS    Genoscope.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .997
                     /organism "Tetraodon nigroviridis"
                     /db_xref "taxon:99883"
                     /clone="042E03"
                     /clone_lib="A"
                     /note "Genoscope sequence ID : C0AA042A03Z13 end : 17"
BASE COUNT      302 a    209 c    208 g    265 t     13 others
ORIGIN


alignment_scores:
            Quality:  304.50              Length:    243
              Ratio:  3.460                 Gaps:      2
 Percent Similarity:  86.214   Percent Identity: 82.099

alignment_block:
 US-09-126-945B-2 x CNS05KTR   ..

 Align seq 1/1  to: CNS05KTR  from: 1  to: 997

      241 AspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGl 287
          ::|||||||||||||    |||||||||||||||||    |||:::||
        1 GAATCTTCCTGCTCCGGGNACCCATCCATCTGTGCCAGTTNCTCCCACA  50

      287 nLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnL 274
          ||||   :::||||||||||:::|||||||||   |||||||||||||||
       51 ACTCTTNTTTAAACCCCACAACTACGGGCGCTGCATCCGCTGGCTGAACA 100

      274 ysGlnLysGlyIle..................................... 278
          ||:::|||||||||
      101 AGGACAAAGGTATCTTCACCTGTTTCCTGCACGTTTACATTTTAAANATA 150

      278 ..... 278

      151 GATGTTCTAACAGTGTGCCTGCAGCGAGGAACTGATCTGCACCACCACTG 200

      278 ..... 278

      201 CAGATTAAACAATGAAAGAAAACATCTCAAATGCATGTTTCTTTTATTAA 250

      278 ..... 278

      251 ATTCCCATATTTATTAAGACACTGTATACATTTGAAAAAAAAAAAAAAII 300

      278 ..... 278

      301 AGTTGTGGIAAATATGACTGTTGTTTGATATAATTCTTAGTATAAATATT 350

      278 ..... 278

      351 CATTIAAACGAGATICTTICAAACTGTAGTATCATAGAAACTCCAAAACA 400

      278 ..... 278

      401 GATGTGTTTTGTCATTTPGTTTGCGTTTTTTTAAGAGTACGTCTGAAATT 450

      278 ..... 278

      451 CTAGTTTAAATAAAACGAAAACCTCAGCGAAAGACATGAAAGACATCATC 500

      278 ..... 278

      501 CTGGAACTTGTTTCACTGTTAGCAAAAACCCACTGTCTTATTGTTCCTTC 550

      279 .....Phe..LysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyI 293
                ||| ||||||||||||||||||:::||||||||||||||||||||:
      551 AGGTATTTTTAAAATTGAGGACTCGGCTCACGTGGCCAGGCTTTGGGGCC 600

      293 leArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIle 309
          ::|||||||||||||||||||||||||    :::||||||||||||||||
      601 TCCGGAAGAATCGACCTGCCATGAACTANGAAAAGCTGAGCGCTCCGAAA 650

      310 ArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnAr 326
          ||||||   ||||||||||||||||||||||||||||||:::|||||||
      651 CGTCAGAC.TACAAGAAGGGCATCATTCGCAAGCCCGATGTGTCCCAGAG 699

      326 qLeuValTyrGlnPheValHisProIle 335
          |||||||||||||||||||||||||:::
      700 ACTGGTCTACCAGTTCGTGCACCCAGIG 727

seq_name: qb_est99:BG328411

seq_documentation_block:
LOCUS       BG328411       899 bp    mRNA             EST       27-FEB-2001
DEFINITION  602427749F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546860 5',
            mRNA sequence.
ACCESSION   BG328411
VERSION     BG328411.1  GI:13134849
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 899)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1234  row: b  column: 13
            High quality sequence stop: 602.
FEATURES             Location/Qualifiers
     source          1. .899
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:4546860"
                     /clone_lib "NIH_MGC_15"
                     /tissue_type "adenocarcinoma cell line"
                     /lab_host "DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      163 a    346 c    277 g    113 t
ORIGIN


alignment_scores:
            Quality:  308.50              Length:    121
              Ratio:  3.005                 Gaps:      6
 Percent Similarity:  83.471   Percent Identity: 71.901

alignment_block:
 US-09-126-945B-2 x BG328411   ..

 Align seq 1/1  to: BG328411  from: 1  to: 899

        1 MetGlySerAlaSerProGly.LeuSerSerValSerProSerHisLeuL 17
```

```
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||
      412 ATGGGCAGCCCCAGCCCGGGCTCTGAGCAGCGTATCCCCCAGCCACCTCC 461

       17 enLeuProProAspThrValSerArgThrGlyLeuGluLys:AlaAlaAl 33
          |||||||||||||||||   ||||||||||||||||||:: ||||||||
      462 TGCTGCCCCCGACACGGG.TCGCGGACAGGCTGGAGAACCGCGGCAGC 510

       33 aGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaI 50
          |||||||:::|||||||||||||||||||||||||||||||||||||  |
      511 GGGGGCAGCGGCTCTCGAGAGACGGGACTGGAGTCCCAGTCCACCCGCA 559

       50 hrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeu 66
          ||||:::::::      :::||||||||||||||||||   |||||||||
      560 CGCCGAGCAGGCCT...GCCGCCTTCTACCTCTCCTACTT.GACATGCTG 605

       67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 83
          ||||||||||||      |||   :::|||||||   :::   ::::::
      606 TACCCTGAGGACCCAGCCTGGCACGGCAAGGCCCTGGCGCCAGCAGCCGC 655

       84 qGluGluProProGln.GluProGlnGlnCysProValIleAspSerGln 99
          :::||||||||||||| ||||||||||||||||||          :::
      656 GAAGGAGCCACCTGAGCGAGCCTGACGAGTGCCCGCCATG....ACAGCC 701

      100 AlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGl 116
          |||:::|||||||||||||||||||||      |||   ||||||||
      702 AAGCCAGGGGGCAGCCTGGACTTGGTGCCG...CGGCTAGACCTGGAGGA 748

      116 qHisSer 118
          ||||:::
      749 GCACGCT 755

seq_name: qb qss5:CNS04XXT

seq_documentation_block:
LOCUS        CNS04XXT      1071 bp    DNA              GSS        26 JUL 2000
DEFINITION   Tetraodon nigroviridis genome survey sequence T7 end of clone
             027A24 of library A from Tetraodon nigroviridis, genomic survey
             sequence.
ACCESSION    AL312266
VERSION      AL312266.1  GI:9545134
KEYWORDS     GSS; genome survey sequence.
SOURCE       Tetraodon nigroviridis.
  ORGANISM   Tetraodon nigroviridis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
             Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
             Tetraodontidae; Tetraodon.
REFERENCE    1  (bases 1 to 1071)
  AUTHORS    Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fischer,C.,
             Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
             Saurin,W. and Weissenbach,J.
  TITLE      Estimate of human gene number provided by genome wide analysis
             using Tetraodon nigroviridis DNA sequence
  JOURNAL    Nat. Genet. 25 (2), 235 238 (2000)
  MEDLINE    20296633
REFERENCE    2  (bases 1 to 1071)
  AUTHORS    Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
             Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
             Bernot,A. and Weissenbach,J.
  TITLE      Characterization and repeat analysis of the compact genome of the
             freshwater pufferfish tetraodon nigroviridis
  JOURNAL    Genome Res. 10 (7), 939 949 (2000)
  MEDLINE    20459837
REFERENCE    3  (bases 1 to 1071)
  AUTHORS    Genoscope.
  TITLE      Direct Submission
  JOURNAL    Submitted (12 APR 2000) to the EMBL/GenBank/DDBJ databases
COMMENT      This sequence is a single read and was generated as part of a large
             scale clone end sequencing project of the Tetraodon nigroviridis
             genome. For more information, please take a look at
             http://www.genoscope.cns.fr/tetra.fa
FEATURES             Location/Qualifiers
     source          1..1071
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="027A24"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA027BA12C1 end : T7"
BASE COUNT       194 a    150 c    162 g    483 t     82 others
ORIGIN


alignment_scores:
            Quality:  292.00              Length:      59
              Ratio:  5.034                 Gaps:       0
 Percent Similarity:  98.305   Percent Identity: 91.525

alignment_block:
 US-09-126-945B-2 x CNS04XXT   ..

 Align_seq 1/1  to: CNS04XXT  from: 1  to: 1071

      277 GlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIl 293
          |||||||||  ||||||||||||||:::|||||||||||||||||||||::
      559 GGTATTTTTNANATTGAGGACTCGGCTCACGTGGCCAGGCTTGGGGCCT 608

      293 eArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleA 310
          :|||||||||||||||||||||||||||||||||||||||||||||||||
      609 CCGGAAGAATCGACCTGCCATGAACTACGACAAGCTGAGCCGCTCGATAC 658

      310 rgGluTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArg 326
          ||||||||||||||||||||||||||||||||||||||:::|||||||||
      659 GTCAGTACTACAAGAAGGGCATCATCGCAAGCCCGATCTGTCCCAGAGA 708

      327 LeuValTyrGluPheValHisProIle 335
          |||||||||||||||||||||||||:::
      709 CTGGTCTACGAGTTCGTGCACCCAGTG 735

seq_name: qb qss26:AZ025433

seq_documentation_block:
LOCUS        AZ025433      414 bp    DNA              GSS        25 FEB 2000
DEFINITION   RPCI-23-345K9.1V RPCI-23 Mus musculus genomic clone RPCI-23-345K9,
             DNA sequence.
ACCESSION    AZ025433
VERSION      AZ025433.1  GI:7100817
KEYWORDS     GSS.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 414)
  AUTHORS    Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinrot
             ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
             and Fraser,C.M.
  TITLE      Mouse BAC End Sequences from Library RPCI-23
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Shaying Zhao
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301-838-0200
             Fax: 301-838-0208
             Email: szhao@tigr.org
             Clones are derived from the mouse BAC library RPCI-23. For BAC
             library availability, please contact Pieter de Jong
             (pieter@dejong.med.buffalo.edu). Clones may be purchased from
             BACPAC Resources (http://bacpac.med.buffalo.edu/orderinfoframe.htm)
             or from Research Genetics (info@resgen.com). BAC end pairs:
             http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
             Plate: 345  row: K  column: 9
             Seq primer: T7
             Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..414
                     /organism="Mus musculus"
```

```
                     /strain "C57BL/6J"
                     /db_xref "taxon:10090"
                     /clone "RPCI-23-345K9"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note "Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoKI and EcoKI Methylase.  Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT        79 a    102 c    129 g    104 t
ORIGIN


alignment_scores:
            Quality:  280.00              Length:      54
              Ratio:  5.185                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 96.296

alignment_block:
 US-09-126-945B-2 x AI025433/rev  ..

 Align seq 1/1  to reverse of: AI025433  from: 1  to: 414

      282 GluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgPr 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      385 GAGGACTCAGCACAGGTGGCCCGACTGTGGGGTGTGCGCAAGAACCGGGC 336

      298 oAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyrTyrLysL 415
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      335 AGCCATGAACTATGATAAACTAAGCCGCTCCATCCGCGAGTATTACAAGA 286

      315 yaGlyIleIleAspLysProAspIleSerGluArgLeValTyrGluLeLy 414
          ||||||| ||||||||||||||||||||||||||||||||||||||||||
      285 ACGCCATCATTCGTAAACCCGACATCTCTCAGCGCCTTGTCTACCAATTT 236

      332 ValHisProIle 335
          ||||||||||:::
      235 GTGCATCCAGTC 224

seq_name: qb_est20:AI451780

seq_documentation_block:
LOCUS        AI451780       544 bp    mRNA            EST       15-MAR-2000
DEFINITION  xb22c07.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:330181 5', mRNA sequence.
ACCESSION    AI451780
VERSION      AI451780.1  GI:4306393
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 544)
  AUTHORS    Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
             Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
             ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,K., Ritter
             ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,K.,
             Waterston,R. and Wilson,R
  TITLE      The WashU-NCI Mouse EST Project 1999
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Marra M/WashU-NCI Mouse EST Project 1999
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 414 286 1800
             Fax: 414 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             This read is a FEASEQUENCE of a previously sequenced mouse clone
             This read has been verified (found to hit its original sequence in the
             correct orientation)
             MGI:211581
             Seq primer:  40RP from Gibco
             High quality sequence stop: 462
             POLYA=No.
FEATURES             Location/Qualifiers
     source          1..544
                     /organism="Mus musculus"
                     /db_xref "taxon:10090"
                     /clone="IMAGE:330181"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage "19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note "Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCAIIIIIIIIIIIIIIII 3'],
                     double stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot  5. Library constructed by Benito
                     Soares and M.Fatima Bonaldo.  RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT       104 a    190 c    153 g     97 t
ORIGIN


alignment_scores:
            Quality:  278.50              Length:      72
              Ratio:  4.220                 Gaps:       1
 Percent Similarity:  91.667   Percent Identity: 76.389

alignment_block:
 US-09-126-945B-2 x AI451780  ..

 Align seq 1/1  to: AI451780  from: 1  to: 544

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
          ||||||||||||||||||||||||||||:::|||||||||||:::  |||||
      331 ATGGGCAGTGCCAGCCCAGGCCTGAGCAACGTGTCCCCCGGTCACCTGCT 380

       17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 34
          ||||  |||||||||:::  ||||||||  |||||||||||||:::|
      381 ACTGTTCCCAGATGTGCACCACGAACAGGGACGGAAGGCAGCATCAG 430

       34 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
          |||||::|||  |||:::::::::::::::::::::::::::::::::
      431 GAGCAATGGGCCCTGAGAAGCAGGAATGGAGTCCTAGTCCACCCGCCACC 480

       51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
          ||||||||||||||||||||||||||||||||:|||||||:::|||  ||
      481 CCTGAGCAGGGCCTGTCTGCTTTCTACCTCTCTTACTTTAACATG...TA 527

       67 rProGluAspSerGly 72
          ||||:::||||||||||
      528 TCCCGACGATAGCAGC 543

seq_name: qb_est54:AW988753

seq_documentation_block:
LOCUS        AW988753       618 bp    mRNA            EST       02 JUN 2000
DEFINITION  ut20107.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1511941 5' similar to TR:O95238 O95238 EIS TRANSCRIPTION
            FACTOR PDEF. ;, mRNA sequence.
ACCESSION    AW988753
VERSION      AW988753.1  GI:8183979
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 618)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:9988793
            Seq primer:  40RP from Gibco
            High quality sequence stop: 407.
FEATURES             Location/Qualifiers
     source          1..618
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1511941"
                     /clone_lib="Soares_mammary_gland_NMLM1"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I   oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       114 a    225 c    164 g    115 t
ORIGIN


alignment scores:
            Quality:  268.50              Length:     73
              Ratio:  4.141                 Gaps:     1
 Percent Similarity:  89.041   Percent Identity: 74.973

alignment_block:
 DS 09 126 945B 2 x AW988753  ..

 Align seq 1/1  to: AW988753  from: 1  to: 618

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
          ||||||||||||||||||||||||||||||:|||||||||:  |||||
      403 ATGGGCAGTGCCAGCCCAGGCCTGAGCAACGTGTCCCCCGGTTGCCTGCT 452

       17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 44
          ||||  |||||||:::::  |||||||||  |||||||||||||:::
      453 ACTGTTCCCAGATGTGGCACCACGAACAGGCAGCGAGAAGGCAGCATCAT 502

       44 lyAlaValGlyLeuGluArgArgAspTrpSerProSerPr:Pr:Alath: 5:
          |||:::|||  |||:::::::::|||||||||||||||||||||||||
      503 GAGCAATGGGCCCTGAGAAGCAGGAATGGAGTCCTAGTCCACCCGCCACC 552

       51 ProGlnGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
          |||||||||||||||||||||||||||||||||  |||:::|||  ||
      553 CCTGAGCAGGGCCTGTCTGCTTTCTACTCTCTTAGTTTAACATG...TA 599

       67 rProGluAspSerSerTrp 73
          ||||:::|||||||||||||
      600 TCCCGACGATAGCAGCTGG 618

seq name: gb htc:AK014602

seq documentation block:
LOCUS       AK014602      4480 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched
            library, clone:4632432C17, full insert sequence.
ACCESSION   AK014602
VERSION     AK014602.1  GI:12852559
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 0 day neonate skin cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:4632432C17.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499474
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system -384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   the RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 4480)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,I., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Nozaki,F., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,R., Saito,A., Sakii,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shimagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,I., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAAGGAICCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full length by cap trapper. cDNA went
            through one round of normalization to Rot  10.0 and subtraction to
            Rot  100.0. Second strand cDNA was prepared with the primer
            adaptor of sequence [5'
            GAGAGAGAGATCTCGAGTIAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. cDNA of size comprised between 4 and 7 kb was
            selected before cloning. Vector: a modified pBluescript KS(-) after
            bulk excision from lambda ZAP-I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES             Location/Qualifiers
     source          1..4480
```

```
                     /organism "Mus musculus"
                     /strain "C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1897048"
                     /db_xref "MGD:MGI:2145456"
                     /clone="4632432C17"
                     /tissue_type "skin"
                     /clone_lib "RIKEN full-length enriched mouse cDNA library"
                     /dev_stage "0 day neonate"
     CDS             163..1569
                     /note "putative"
                     /codon_start 1
                     /protein_id "BAB29457.1"
                     /db_xref "GI:12852560"
                     /translation="MNDPCLKNMLQVAPVANSFPGTLRKVLDAFLEEWSLPAVLDSLS
                     EDQILQFVPTGLDSVSHDSASCFLPLTLPCSKAVMSQAVKAALSCPLRPQRRLCLPKN
                     PWLWSEQQVCQWLLWATNEFSLVNVNLHQFGMNGQMLCNLGKEKFLELAPLFVGDILW
                     EHLEQMIKENQEKIRQYEENSHLNAVPRWINSNTLTCPMLAEYCMQAFNYLRNLL
                     DSMCPPSATPAALGSELQMLPFSPLNTVNVNYQSISQQLTSSNVNLLNNNSCKPPGHR
                     SPENGGDSFESSDSLLRSWNSQSSLLDVQRVPSFESFEEDCSQSLCLSKLIMSFKDYI
                     QERSDPVEQGKPVTPAAVLAGFTGSCPIQLWQFLLELLSDPSCQSFISWTGDQWEPKL
                     ADPDEVARRWGKFKNKFFMNYEKLSRGLRYYYDRNIIARKLSQRSYVYRFVCDIQNLLG
                     FIPEELHAILGVQPDTED"
     polyA_signal    3455..3460
                     /note "putative"
     polyA_site      3480
                     /note "putative"
BASE COUNT      826 a    915 c    846 g    893 t
ORIGIN


alignment_scores:
            Quality:  264.50              Length:     501
              Ratio:  1.242                 Gaps:      20
 Percent Similarity:  42.515   Percent Identity:  23.154

alignment_block:
 US-09-126-945B-2 x AK014602   ..

 Align Seq 1/1  to: AK014602  from: 1  to: 3480

       8 LeuSerSerValSerProSerHisLeuLeuLeuProProAspThrValSe  24
         |||      ||||||||||   ||||||      |||              ||
      17 CTGAGGCCCGTCAGTCCCCGCCACCTTCCCCCGGCCCCGCGCCCCGGATC  66

      24 rArgThr.GlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArg  40
         |      ||||||      ::::::  |||  :::      ::::||||
      67 GGCCCTACGGCCTC......GTCTCGCCCGGCCTTGCGGCGCCGGGACCGC 110

      41 ArgAsp...TrpSerPro.SerProProAlaThrProGlu.....GlnG  54
         ||||||      ||||||    ||||||||||   |||          :
     111 CGCGATCTCCTCTCCCCGCCGCCCTCCGGCTGGCCCTGCCTCGCTGCGGCG 160

      54 lyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAsp  74
         ::::::::   |||   ::::::   ::::||||||||   |||
     161 CGATGAATGACTTTGGAATCAAGAACATGGACCAAGTGGCCCCTGTCGCC 210

      71 SerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluProPr  87
         :::|||:::                ||||:   :::|||:::
     211 AACAGTTTTCGT............GGGACACTCAAGCGCCAGCCAGACCCT 248

      87 oGlu......GluProGluGlnCysProValIleAspSerGlnAlaProA 102
           :::    :::  :::      ||||::   |||   :::
     249 TGACACCTTCGATGGCTCTCTGTTTGCTGTGCTCCCTTCTCTCAGTGAAG 298

     102 laGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSer 118
           :::|||:::   ||||||   ||||||          ::::::::
     299 ATCAGACACTCCAAGAAGTGCCCACGGGCCTGGATTCTGTCTCCCATGAC 348

     119 ........................LeuGluGlnValGlnSerMetValVa 137
                                 |||          ::::::   ||||:
     349 TCGGCCAGCTGCGAGCTGCCTTTGCTCACTCCCTGCAGCAAGGCAGTGAT 398

     127 lGlyGluValLeuLys.............AspIleGluThrAlaCysLysL 140
         ::::::::::::|||||                ::::::          :::
     399 GAGCCAAGCCTTAAAAGCCACCTTCAGTGGCTTCCAAAAGGAGCAACGAC 448

     140 euLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGln 156
           ||||:::|||      :::|||     ||||||   :::::::|||
     449 GTCTTGGCAGCCCAAAACCCTGGCTCTGGAGTGAGCAGCAGGTGTGC 498

     157 LysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLy 173
         :::||||||.||||||::   :::::::::   |||      ::::::
     499 CAGTGGCTTCTCTGGGCCACCAACGAGTTCAGCCTGGTGAATGTGAACCT 548

     173 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluG 190
                   :::   ||::::   ||||||   ::::::::|||:
     549 GCACCAGTTTGGCATGAACGGCCAGATGCTGTGTAACCTCGGCAAGGAGC 598

     190 lnPheArgGlnArgSerPro...LeuGlyGlyAspValLeuHisAlaHis 205
         ::|||   :::   ::|||   :::   ||||||::|||      |||
     599 GCTTCCTGGAGCTGGCGCCTGACTTTGTGGGTGACATCCTCTGGGAACAT 648

     206 Leu ..........                    AspIleTrpTy 210
         |||                              |||   :::::
     649 CTAGAGCAGATGATCAAAGAGAACCAAGAAAAGACAGAAGACCAATATGA 698

     210 sSerAlaAla.....................TrpMetLysGlnArgThr. 229
         ::::   :::                   |||:::::::::::::|||
     699 GGAAAACTCTCACCTCAACGCGGTTCCTCATTGGATCAACAGCAATACAT 748

     229 ..................................................... 229

     749 TAGGCTTCAGCATGGAACAGGCTCCATATGGAATGCAGGCACCAAACTAC 798

     220 .....................................SerProGl 232
                                             :::|||::
     799 CCCAAAGACAATCTCCTGGACAGCATGTGCCCGCCATCGGCCACGCTGC 848

     232 yAla..................................................  233
         :|||
     849 AGCTCTGGGCTCTGAGCTCAAGATGTTGCCAAGTCTCGGCTCAACACCG 898

     224 .....IleHisTyrCysAlaSerThrSerGlu..................  242
          ::::::|||||:::   ::::::::                       
     899 TCAATGTCAATTACTGTTCCATCAGCCAGGACTTCCCCAGCAGCAACGTG 948

     242 ..................................................... 242

     949 AATTTGCTCAACAACAATTCTGGAAAACCCAAGGACCACGACTCTCCAGA 998

     233 .....................................GluSerTrpI 246
                                             :::||||||:
     999 GAACGGTGGGGACAGCTTCGAGAGCTCCGACTCGCTGCTGAGGTCCTGGA 1048

     236 hrAsp.....................................SerGlu 249
         :::::                                   ::::::
    1049 ACAGCCAGTCGTCCCTACTGGATGTACAGCGGGTACCTTCCTTCGACAGC 1098

     240 ValAspSerSerCysSer.............................. 245
           ::::::::::|||||
    1099 TTTGACGAGGACTGTAGCCAGTCTCTGTGCCTCAGTAAGCTGACCATGTC 1148

     245 ..................................................... 245

    1149 CTTCAAGGACTACATCCAAGAGAGGAGCGACCCAGTCGAGCAAGGCAAAC 1198

     246 .....................................GlyGlnProIle 259
                                             :::   ||||||
    1199 CAGTTATTCCTGCAGCAGACTGGCTGGCTTCACTGGAAGCGGACCAATC 1248

     250 HisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGl 266
         :::|||||||||||||||   |||   ||||||:::      |||
    1249 CAGTTGTGGCAGTTTCTTCTGGAG...CTACTCTCTGACAAGTCCTGTCA 1295

     266 yArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluA 283
```

```
            ||||||   |||   :::   :::          |||||||:::   |
    1296 ATCTTTCATCAGCTGGACGGGG...GATGGATGGGAGTTCAAGCTTGCTTG 1342

     288 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299
         ||       :::||||||||||   ||||||   ||||||||||:::|||
    1343 ACCCCGATGAGGTTGCCCGCCGGTGGGGGAAGAGGAAAAATAAACCAAAG 1392

     300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
         ||||||||||:::||||||||||||||:::::|||   ||||||   ||:::
    1393 ATGAACTACGACAAGCTGAGCCGGGGCTTACGTTACTACTACGACAAGAA 1442

     316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal 332
         :|||||:::|||   :::   :::::::|||   |||||||:::||||||
    1443 CATCATCCACAAGACTTCG...GGCAAGCGCTACGTGTACCGTTTCGTA 1488

Seq_name: qb_est90:BF589055

seq_documentation_block:
LOCUS       BF589055      640 bp    mRNA            EST       12-DEC-2000
DEFINITION  nac29d04.x1 Lupski_sciatic_nerve Homo sapiens cDNA clone
            IMAGE:3894614 3' similar to TR:O95238 O95238 ETS TRANSCRIPTION
            FACTOR PDEF. ;, mRNA sequence.
ACCESSION   BF589055
VERSION     BF589055.1  GI:11681379
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lupski
             cDNA Library Preparation: Lupski Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            info@image.llnl.gov
            Seq primer:  40UP from Gibco
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1..640
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3894614"
                     /clone_lib="Lupski_sciatic_nerve"
                     /sex="male"
                     /tissue_type="sciatic nerve"
                     /dev_stage="adult, 70 yr"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT (Life Technologies); Site_1:
                     NotI; Site_2: SalI; cDNA made by oligo dT priming;
                     directionally cloned using the following adaptors:
                     5' TCGACCCACGCGTCCG 3' and
                     5' GACTAGTTCTAGATCGCGAGCGGCCGCCCT(T5)-3'; Size selected -
                     1 Kb for average insert length 1.47 kb. This is a primary
                     library, non-amplified. Library constructed by Life
                     Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
                     College of Medicine) and is available through Life
                     Technologies."
BASE COUNT       103 a    182 c    218 g    136 t      1 others
ORIGIN


alignment_scores:
             Quality:  264.00             Length:      54
               Ratio:  5.058                Gaps:       0
 Percent Similarity:  96.296   Percent Identity: 94.444

alignment_block:
 US-09-126-945b-2 x BF589055/rev    ..

 Align seq 1/1  to reverse of: BF589055  from: 1  to: 640

      282 GluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgPr 298
          ||||||||||||:::|||||||||||||||||||||   ||||||||||
      640 GAGGACTCAGCCAGGGTGGCCCGACTGTGGGGCATCNGCAAGAACGGTCC 591

      298 nAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysL 315
          ||||||||||||||||||||||||||||||||||||||||||||||   ||||
      590 CGCCATGAACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAAGA 541

      315 ysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      540 AGGGCATCATCCGGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTC 491

      332 ValHisProIle 435
          ||||||||||||
      490 GTGCACCCCATC 479

Seq_name: qb_qss3:CNS027SY

seq_documentation_block:
LOCUS       CNS027SY      882 bp    DNA             GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence 17 end of clone
            243D23 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL185083
VERSION     AL185083.1  GI:7823187
KEYWORDS    GSS: genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 882)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 882)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..882
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="243D23"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : COAG243CB12A1 end : 17"
BASE COUNT       210 a    227 c    215 g    222 t      8 others
ORIGIN


alignment_scores:
             Quality:  257.00             Length:      95
```

```
            Ratio:  3.671                Gaps:      3
 Percent Similarity:  73.684   Percent Identity: 55.789

alignment_block:
 US-09-126-945B-2 x CNS027SY   ..

 Align seq 1/1  to: CNS027SY  from: 1  to: 882

     144 ThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrpLeuLe 193
         :::::!!!!!!:::!!!!!!!!!!      :::!!!!!!!!!!!!!!!!!!
     447 TCCTCAGACCCCTTGGACTGGAGCYGTGTGCAGCTTCAGAAGTGGCTCCT 496

     160 uTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnG 177
         !!!!!!!!!!!   !!!!!!!!!!!   ::::::      !!!!!!!
     497 CTGGACGGAGCACCTGTACAGGCTGCCACAGGTCAGCACCATGTTCCAGG 546

     177 luLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGln 193
         !!!!!:::!!!::::::!!!!!!:::!!!:::!!!   :::!!!!!!!!!
     547 AGCTGACGGCAGGGACCTGTGCTCCATGACGGAGGCGGACTTCAGACAG 596

     194 ArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLy 210
         !!!!!!      !!!!!!!!!!!!!:::!!!!!!!!!!!!!!!!!::
     597 CGCTCCTCGCAGTTCGGAGATGTCCTCTACGCCCACCTGGACATCTGGAG 646

     210 sSerAla.....AlaTrpMetLysGluArgThrS 220
         :!!!::::              :::!!!!!!!!!
     647 ATCCAGTAAGAACCGAAGAATTGAATAAGAGGTGGTTAAAGAAAGGTTT. 695

     220 erProGlyAlaIleHisTyrCysAlaSerThrSer 241
           !!!!!!!::         !!!!!!!!!!!!
     696 .....GGAGCTTTA.........GCCTCCACCTCC 716

seq_name: qb_est97:BG172401

seq_documentation_block:
LOCUS       BG172401      966 bp    mRNA            EST       06-FEB-2001
DEFINITION  602333568F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4456745 5',
            mRNA sequence.
ACCESSION   BG172401
VERSION     BG172401.1  GI:12679104
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 966)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10252  row: q  column: 18
            High quality sequence stop: 653.
FEATURES             Location/Qualifiers
     source          1..966
                     /organism "Mus musculus"
                     /strain "FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4456745"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="3 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ, mammary, Vector: pCMV-SPORT6, Site_1: SalI,
                     Site_2: NotI; Cloned unidirectionally.  Primer: Alias dI.
                     Library constructed by Life Technologies. Investigator
                                providing samples: Gilbert Smith, NIH"
BASE COUNT      265 a    245 c    268 g    188 t
ORIGIN


alignment_scores:
          Quality:  249.00              Length:    198
            Ratio:  2.024                 Gaps:      4
 Percent Similarity:  62.121   Percent Identity: 51.818

alignment_block:
 US-09-126-945B-2 x BG172401   ..

 Align seq 1/1  to: BG172401  from: 1  to: 966

     135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
         :::!!!!!!!!!                 ::::::   !!!      !!!::
      62 CAGACAGCCTGTGATTCCTACTGGACATCAGTGCACCCTGAATACTGGAC 111

     151 rProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 168
         :      :::!!!   :::!!!!!!   :::          !!!!!!!!!
     112 CAAGCGCCATGTCTGGGAATGGCTCCAATTCTGCTGCGACCAGTACAAGC 161

     168 euProProMetGlyLysAlaPhe......GlnGluLeuAlaGlyLysGlu 182
         !!          :::!!!          :::!!!!!!!!!!   :::
     162 TTGATGCCAACTGCATCTCCTTCTGTCACTTCAACATCAGCGGCCTGCAG 211

     183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199
         !!!!!:::!!!::::::!!!:::!!!   :::   :::   :::   !!
     212 CTCTGCAGCATGAGCAGGAGGAGTTCATTGAGCCAGCCGGCATCTGTGG 261

     199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMet. 215
         !:::   !!:::       !!!:::       :::!!!   ::::::
     262 GGAGTACCTGTACTTCATTCTCCAGAACATTCGCTCGCAAGGTTACTCCT 311

     216 ............LysGluArgThrSerProGlyAlaIleHistyr...... 226
                     !!!:::!!:::   !!!         !!!
     312 TTTCAATGATGCTGAAGAGACCAAGACTGGCATCAAAGACTATGCTGAT 361

     227 ......CysAlaSerThrSerGluGluSerTrpThrAspSerGlnValAs 241
               !!!   :::!!!!!!      :::      ::::::::
     362 TCCAGTTGCTTGAAAACAAGT......GGCATCAAGAGTCAAGACTGTCA 405

     241 pSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGlnL 258
          !!!      !!!   !!!      !!!!!!!!!:::!!!:::::::::!
     406 CAGCCGAACAAGCCTCCAAAGTTCTCACCTGTGGGAATTTGTCAGAGACT 455

     258 euLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 274
         !!!!!!!:::!!!:::::        ::::::::!!!   ::::::
     456 TGCTGCTGTCCCCTGAAGAGAACTGTGGCATCCTGGAATGGGAAGACAGG 505

     275 GluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTr 291
         !!!:::!!!!!!!!!:::::      !!!      :::!!!::::::!!
     506 GAGCAGGGCATTTTCGAGTGGTTAAGTCAGAAGCCCTGGCAAACATGTG 555

     291 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 308
         !!!!.  !!!   ::::::      !!!:::!!!:::!!!!!!!!!!!!
     556 GGGACAAACGACGAACAATCACAGGATGACGTACGACAAGCTGAGCCGAG 604

     308 erIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAsp 322
           :::!!!   !!!!!!!!!   !!!!!!:::::::::   !!!
     605 CCTGAGATACTACTATAAAACGGGAATTCTGGAGCGGCTTGAC 648

seq_name: qb_est89:BF536841

seq_documentation_block:
LOCUS       BF536841      629 bp    mRNA            EST       11-DEC-2000
DEFINITION  602048419F1 NCI_CGAP_Sk2 Mus musculus cDNA clone IMAGE:4187884 5',
            mRNA sequence.
ACCESSION   BF536841
VERSION     BF536841.1  GI:11624199
KEYWORDS    EST.
SOURCE      house mouse.
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9611  row: e  column: 05
            High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1..629
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4187884"
                     /clone_lib="NCI_CGAP_Sc2"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: cell; Vector: pSPORT1; Site 1:
                     NotI; Site 2: Call; Cloned unidirectionally, Primer: oligo
                     dT, Average insert size 1.4 kb. Constructed by Life
                     Technologies. Note: this is a NCI CGAP library."
BASE COUNT       118 a    224 c    178 g    106 t      3 others
ORIGIN


alignment scores:
            Quality:  242.50            Length:     92
              Ratio:  3.191               Gaps:      2
 Percent Similarity:  82.609   Percent Identity: 64.130

alignment block:
 US-09-126-945B-2 x BF536831  ..

 Align seq 1/1  to: BF536831  from: 1  to: 629

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSer.....Hi 15
          ||||||||||||||||||||||||||::||||||||||:::
      458 ATGGGCAGTGCCAGCCCAGGCCTGAGCAACGTGTCCCCCNCGTTGCCTTG 407

       15 sLeuLeuLeuProProAspThrValSerArgThrGlyLeuGluAlaA 62
                 :::|||          ::::::::          ||||||||||
      408 CTACTGGTTCCCAGNATGTGGCACCACGAAACAGGNGACGGAGAAGGCAG 457

       32 laAlaGlyAlaValGlyLeuGluArgArgAspProSerProSerProPro 80
          ||:::|||||::|||   |||::::::::::||||||||||||||||||
      458 CATCAGGAGCAATGGGCCCTGAGAAGCAGGAATGGAGTCCTAGTCCACCC 507

       49 AlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspPh 97
          ||||||||||||||||||||||||||||||||||||||||||||:::||
      508 GCCACCCCTGAGCAGGGCCTGTCTGCTTTCTACTCTCTTACTTTAACAT 556

       65 tLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSer: 82
          |   ||||||:::|||||||||||||:::||||||:::|||   |||   :
      557 G...TATCCCGACGATAGCAGCTGGGTTGCCAAAGCCCCCGAGGCCCCTC 603

       82 erArgGluGluProProGluGluPro 90
          ::   |||   |||||||||||||||
      604 CCGGGGAGGA.CCACCCGAGGAGCCC 628

Seq name: gb_htc:AK018368

Seq documentation block:
LOCUS       AK018368     2478 bp    mRNA            HTC       08-FEB-2001
DEFINITION  Mus musculus 16 days embryo lung cDNA, RIKEN full length enriched
            library, clone:B430401F14, full insert sequence.
ACCESSION   AK018368
VERSION     AK018368.1  GI:12850036
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (house mouse), 16 days embryo lung cDNA to mRNA.
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:B430401F14.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499474
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Ikeda,R., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwara,Y., Hara,E., Ishida,J., Ishiwa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384 format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20540913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 2478)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,I., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,I.,
            Miyazaki,A., Nishi,K., Nomura,K., Nomazaki,S., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,R., Saito,H., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,I.,
            Soeabe,Y., Suzuki,H., Taqami,M., Taqawa,A., Takahashi,F.,
            Tanaka,T., Tojima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-r@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/)
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full length by cap trapper. cDNA went
            through one round of normalization to Rot 10.0 and subtraction to
            Rot 185.2. Second strand cDNA was prepared with the primer
            adapter of sequence [5'
            GAGAGAGAGATCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3'], cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(-) after
```

```
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SrfI; 3'
            end: BamHI. Host: DH10B.
FEATURES             Location/Qualifiers
     source          1. .2478
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref "taxon:10090"
                     /db_xref "MGD:MGI:1909151"
                     /db_xref "MGD:MGI:1923002"
                     /clone "B430401F14"
                     /tissue_type "lung"
                     /clone_lib "RIKEN full length enriched mouse cDNA library"
                     /dev_stage "16 days embryo"
     CDS             185. .1717
                     /note "putative"
                     /codon_start 1
                     /protein_id="BAB31180.1"
                     /db_xref "GI:12858057"
                     /translation="MEWCFCLQQVPEMVIQESRSEEWRRPLLQEEEEFVDIDLAHDSE
                     ELFQDESQLQFAWLAEAQVPDDDFQFVPDPQSDNCVCHAPPPLKKEEDSPSSEESSC
                     SHEQALGAEVCEEPLYNYCAYDEEDFSGDFEPLTPEATELCPHE_NCLETPPCAILPTS
                     GLTPGAGPVQGVGPAPTPHSLPEPGSQQQTTFAVPRPPHQPLQMPKMMPESQYPSEQRF
                     QRQLSEPSHPFPPQSGVPGDSRPSYHRQMSEPIVFAAPPPLQGFKQEYHDPLYEHGVP
                     GMPGPPAHGFQSPMGIKQEPEDYCADSEVPNCQSSYMRGGYFSSSHEGPPYEKDPRLY
                     FDDICVVPERLEGKVKQEPIMYREGPPYQRRGSLQLWQFLVTLLFDPANAHFIAWTGR
                     GMEFFLVKPERVAPPWGLQLNERAMNYQRLPEADLPLYEICVMEVAERSVVLEVVDG
                     PDALFSMAFPDNQRPPLKAESIGPPLNIEDFTLPLTHFTDNPAYLLDMDRCSSLPYIEGF
                     AY"
BASE COUNT       580 a    705 c    644 g    559 t
ORIGIN


alignment_scores:
            Quality:  239.00              Length:     389
              Ratio:  1.358                 Gaps:      19
 Percent Similarity:  45.244   Percent Identity: 25.964

alignment_block:
 US-09-126-945B-2 x AK018368   ..

 Align seq 1/1  to: AK018368  from: 1  to: 2478

       5 SerProGlyLeuSerSerValSerProSerHis.....LeuLeuLe 18
         :::|||    :::   :::||||||::|||      ||||::
     587 ACCCCGCCCGCCACICCICIGICACCCACCCAICAGAATICCCCIAIIGCC 636

      18 nProProAsp...ThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 54
           |:|||||::   |||:::      :::|||||      :::|||:::
     637 CCCACCCCAGGCAACCCIGCCCACCICAGGICCICACCCCIGGAGCCIGCCC 686

      54 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
         :::   |||:::              :::|||:::||   :::
     687 CAGIICAAGGGGIGGGIGCCI.....GCCCCGACICCCICACICACCI 727

      51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
         |:|||   |||
     728 CCAGAACCIGGA..................................... 739

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
           :::::::::::::::|||:::
     740 .ICACAGCAACAAACATTGCGGGIC.....................? 764

      84 luGluProProGluGluProGluGlnCysProValIle.....AspSer 98
         :::|||||||||||:|||   |||   |||   :::       :::|||
     765 CCCGCCCACCACAICAGCCCCIGCAAATGCCAAAGATGATGCCTGAAAGC 814

      99 GlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGl 115
         |||   ||:::                                    ||
     815 CAGIAICCAICA.....................................GA 828

     115 uGluHisSerLeuGluGlnVal.....................G 129
         |:::::      :::|||:::                     |
     829 ACAGAGAIIICAGCGACAGCIGICIGAGCCCAGCCAICCCIICCCIGCIC 878

     129 InSerMetValValGlyGluValLeuLysAspIleGluThrAlaCysLys 145
         |||||   |||   |||:::          :::   :::          :::
     879 AGICAGGAGIICCIGGAGAIAGICGCCCCAGCIACCACCGGCAGAIGICA 928

     146 LeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGl 196
               :::   |||   |||
     929 GAGCCCAIIGICGCCGCAGCICCGCCACCC..................... 958

     156 nLysTrpLeuLeuTrpThrGluHisGluTyrArgLeuPro.......... 169
              |||          :::::::::|||:::   |||
     959 .......CTTCAGGGATICAAGCAGGAATACCATGACCCACIGIAIGAAC 1001

     170 ......................ProMetGlyLysAlaPheGlnGluLeu 178
                                  |||   :::   :::||||||::
    1002 ArGGGcICCCIGGGArGCcIGcIcc ICCAGCAcAIGGGIICCAGIcACCC 1051

     179 AlaGly..............LysGlnLeuCysAlaMetSerGlnGluGl 190
           |||               ::::::   ||||||   ||||||
    1052 AIGGGGAIc AAGcArGAGCCCCCAGAIIACCIGCIGCIGACICAGAAGIGCC 1101

     190 nPheArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuA 207
              |||   |||   :::   |||       :::
    1102 IAACIGCCAGICAICCTACATGAGAGGCGGGIAIIIC.............. 1138

     207 spIleTrpLysSerAlaAlaArgMetLysGlnArgIhrSerProLysAla 223

    1138 .................................................. 1138

     224 IleHisTyrCysAlaSerThrSerGluGlnSerTrpThrAspSerGlu.. 239
                        |||:::|||:::|||:::::::      ::::::
    1139 ..............TCCAGCAGCCAIGAAGGAIICCCGIAIGAGAAGGA 1178

     240 .............ValAspSerSerCys..................... 244
                        |||:::::|||
    1174 CCCCAGGCIGIACIIIGAIGAIACIIGGCIGGIAICCCGAGAGACIGGAAG 1223

     245 ..................................SerGlyGluPro..... 248
                                          :::|||   |||
    1224 GCAAAGICAAGCAGGAGCCCACCAIGIAICGAGAGGGGCCIICCIIAICAG 1273

     249 .............IleHisLeuTrpGluPheLeuLysGluLeuLeuLeuLy 261
                     :::::|||||||||||||||||      ||||||
    1274 AGACGAGGCICICIGCAGCIGIGGCAGIICCIIGIICACCCICCIGGAIGA 1323

     261 sProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyI 278
         |||   :::   :::::|||||   |||   :::       :::|||:
    1324 CCCA...GCCAAIGCCCACIICAIGGCIGGACIGGC.....CGAGGCA 1364

     278 Ie...PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIle 293
         ::   |||||||::   :::       :::|||||||||   |||||||||
    1365 IGGAAIIIAACCICAIAGAACCGGAAGAGGIIGCICGCCGIIGGGGIAIC 1414

     294 ArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleAr 310
         :::|||||||||||||||||||||||||||||||||||||||||||:||
    1415 CAGAAGAATCGGCCAGCCAIGAACIAIGACAAGCIGAGCCGCICICICCG 1464

     310 qGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGluArgL 327
         |   ||||||:::|||||||||:::::|||       :::::|||
    1465 CIAIIACIAIGAAAAGGGCAICAIGCAGAAGGIG...GCIGGGGAACGCI 1511

     327 euValTyrGlnPheVal 332
           |||:|||::|:|||
    1512 ACGICIACAAAIIIGIG 1528

seq_name: qb_est83:BF133977

seq_documentation_block:
LOCUS       BF133977      747 bp    mRNA            EST       24-OCT-2000
DEFINITION  601778421F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4006732 5',
            mRNA sequence.
ACCESSION   BF133977
VERSION     BF133977.1  GI:10973017
```

```
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9248  row: i  column: 05
            High quality sequence stop: 644.
FEATURES             Location/Qualifiers
     source          1..747
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4000732"
                     /clone_lib="NCI_CGAP_Lu49"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV LTR enhancer; Cloned unidirectionally.  Primer: oligo
                     dT.  Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      193 a    180 c    196 g    168 t
ORIGIN


alignment scores:
            Quality:  237.00              Length:    194
              Ratio:  1.959                 Gaps:      4
 Percent Similarity:  62.371   Percent Identity: 40.412

alignment block:
 US-09-126-945B-2 x BF183977  ..

 Align seq 1/1  to. BF183977  from. 1  to. 747

      185 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpC 234
          :::|||||||||                 ::::::   |||      ||:::
      117 CAGACAGCCTGTGATTCCTACTGGACATCAGTGCACCCTGAATACTGGAC 166

      151 rProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 198
          :      :::|||   :::||||||   :::         ||||||||:|
      167 CAAGCGCCACGTCTGGCAATGGCTCCAATTCTGCTGCGACCAGTACAAGC 216

      168 euProProMetGlyLysAlaPhe.....GlnGluLeuAlaGlyLysGlu 182
          ||           :::|||          ||||||||||||   :::
      217 TTGATGCCAACTGCATCTCCTTCTGCCACTCCAACATCAGCGGCCTGCAG 266

      183 LeuCysAlaMetSerGluGluGlnPheArgGluArgSerProLeuGlyGl 232
          ||||||:::|||:::::::|||:::|||   :::   :::   :::   ||
      267 CTCTGCAGCATGACGCAGGAGGAATTCATTGAGGCAGCCGGCATCTGTGG 316

      233 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMet. 279
          |:::   |||:::      |||:::      :::|||   ::::::
      317 GCAGTACCTGTACTTCATTCTCCAGAACATTCGCTCGCAAGGTTACTCCT 366

      280 ...........LysGluArgThrSerProGlyAlaIleHisTyr..... 300
                     |||:::|||:::   ||        |||
      367 TTTTCAATGATGCTGAAGAAACCAAGACTGGCATCAAAGACTATGCTGAT 416

      301 .....CysAlaSerThrSerGluGluSerTrpThrAspSerGluValAs 341
               |||   :::|||||||      :::      :::::::::
      417 TCCACTTGCTTGAAAACAAGT.....GGCATCAAGACTCAAGACTGTCA 460

      341 pSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGlut 388
          |||      |||   |||       ||||||||||:|||:::::::::|
      461 CAGCCGAACAAGCCTCCAAAGTTCTCACCTGTGGGAATTTGTCAGAGACT 510

      388 ouLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 434
          ||||||||:::|||:::::   :::::::::|||   ::::::
      511 TGCTGCTGTCCCCTGAAGAGAACTGTGGCATCCTGGAATGGGAAGACAGG 560

      435 GluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuIr 484
          |||:::|||||||||:::::      |||      :::|||:::::|||
      561 GAGCAGGGCATTTTCCGACTGGTTAACTCAGAAGCCCTGGCAAAGATGTG 610

      485 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 534
          ||||   ||||||:::::       |||:::|||:::::   :::|||:
      611 GGGACAACGGAACAAGAATGACAGGATGACGTACGAGGAAGCTGAGCGAG 660

      535 erIleArgGlnTyrTyrLysLysGlyIleIle 568
          :::::|||   |||:::|||:::|||||||:::
      661 CCTTGAGATA.TACTTTAAAAGGGGAATTTTG 691

seq name: qb-est95:BF983301

seq documentation block:
ID#DS       BF983301      686 bp    mRNA            EST       24-JAN-2001
DEFINITION  602395527F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4396719 5',
            mRNA sequence.
ACCESSION   BF983301
VERSION     BF983301.1  GI:12386113
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 686)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10096  row: b  column: 16
            High quality sequence stop: 603.
FEATURES             Location/Qualifiers
     source          1..606
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4396719"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo dT primed. Average insert size 1.767 kb. Library
                     enriched for full length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      126 a    192 c    186 g    101 t      1 others
ORIGIN


alignment scores:
            Quality:  234.00              Length:    142
              Ratio:  2.571                 Gaps:      6
 Percent Similarity:  64.085   Percent Identity: 49.437
```

```
alignment_block:
 US-09-126-945B-2 x BF983291   ..

 Align seq 1/1  to: BF983301  from: 1  to: 606

     203 HisAlaHisLeu.................AspIleTrpLysSerAlaAl 213
         ||  ||||||                 :::   |||   ::::::::
      32 CACCAGCACCTCCACCGCCGCCTCGGAACCCAGGCCTGGGGGGCGGTGGC 81

     213 aTrpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerT 230
         :          |||         |||||||||   :::   :::   ::::
      82 GCCGCGTATGGAGCCCCGCGCCCCCGGAGCCTGCCAACATTGCCAACGCCA 131

     230 hrSerGluGluSer..............TrpThrAspSerGluValAsp 241
         |.:::::::   :::              |||   :::
     132 CCGCACGCTACACACAGGTACCCCTGGGATGGCGTGAG.........CAC 172

     242 SerSerCysSerGly.GlnProIleHisLeuTrpGlnPheLeuLysGluL 258
         |:|     ::|||         :::   |||||||||||||||   :::
     173 TCCCCCAGCGATGGACCCATCTGTGACGCTGTGGCAGTTTCTCAGCTGCAG. 221

     258 euLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 274
           ||||||::   ::::::   |||:::::|||   |||   ::::::
     222 ..CTCCTGACAGAGCAAGCCAATGGCCACATCATCTCCTGGACTTCACGG 269

     275 GluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTr 291
         :::   |||   ||||||:::   |||::   :::|||||||||||||
     270 GATGGTGGTGAATTCAAGCTGGTGGATGCAGAGGAGGTGGCCCGGCTGTG 319

     291 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 308
         ||||:::|||||||||:::       |||||||||||||||||||||||:
     320 GGGGCTACGCAAGAACAAGACCAACATGAATTACGACAAGCTCAGCCGGG 369

     308 erIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSer 324
         ::::||,   ||||||   |||:::|||||||||||   :::   :::
     370 CCTTGCGGTACTACTATGACAAGAACATCATCGGCAAGGTCAGC...GGC 416

     325 GlnArgLeuValTyrGlnPheVal 332
         |||::::::|||||::||||||
     417 CAGAAGTTCGTCTACAAGTTTGTG 440

seq_name: qb_est10:AA651760

seq_documentation_block:
LOCUS       AA651760      529 bp    mRNA            EST       13-NOV-1997
DEFINITION  ns32a10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1185306 3'
            similar to qb:M82882 ETS-RELATED TRANSCRIPTION FACTOR ELF-1 (HUMAN
            );, mRNA sequence.
ACCESSION   AA651760
VERSION     AA651760.1  GI:2583412
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   NCI CGAP Lib [...www.ncbi.nlm.nih.gov]
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Unknown library type
            Insert Length: 1189  Std Error: 0.00
            Seq primer: -40m13 fwd. EI from Amersham
            High quality sequence stop: 258.
FEATURES             Location/Qualifiers
     source          1..529
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1185306"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3 Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTT-3'
                     ].  Double stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      187 a    101 c    120 g    121 t
ORIGIN


alignment_scores:
            Quality:  233.50              Length:     86
              Ratio:  3.538                 Gaps:      1
 Percent Similarity:  76.744   Percent Identity: 50.000

alignment_block:
 US-09-126-945B-2 x AA651760   ..

 Align seq 1/1  to: AA651760  from: 1  to: 529

     246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 262
         |||:::   |||:::|||||||:::|||||       ||||||
     126 GGAAACACAATTTATCTTTGGGAGTTTTTACTGGCACTGCTCCAGGACAA 175

     262 oHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 279
           :::     :::::|||:::|||   :::::|||||||||||||
     176 GGCTACTTGTCCTAAATACATCAAGTGGACCCAGCGAGAGAAAGGCATTT 225

     279 heLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLys 295
         |||||:::   ||||||       |||:::|||||||||||   :::|||
     226 TTAAATTGGTGGATTCTAAAGCAGTGTCCAGGTTGTGGGCGAAGCACAAA 275

     296 AsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTy 312
         |||::|||   |||||||||:::   :::::|||:::::|||   ||
     276 AACAAACCTGATATGAATTATGAGACCATCGGAAGAGCACTCAGGTACTA 325

     312 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValT 329
         ||||:::::|||||||:::   |||   :::   :::|||||:|||||
     326 TTACCAAAGGGATATCTGGCAAAACTGGAA...GGTCAGCGCTTGGTGT 372

     329 yrGlnPhe 331
         ||||||||
     373 ATCAGTTT 380

seq_name: qb_est13:AA913571

seq_documentation_block:
LOCUS       AA913571      544 bp    mRNA            EST       10-JUN-1998
DEFINITION  [illegible] [illegible] _MD_1_B7_S1 Homo sapiens cDNA clone
            IMAGE:1525123 3' similar to qb:M82882 ETS-RELATED TRANSCRIPTION
            FACTOR ELF-1 (HUMAN);, mRNA sequence.
ACCESSION   AA913571
VERSION     AA913571.1  GI:3052963
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   NCI CGAP Lib [...www.ncbi.nlm.nih.gov].
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1217   Std Error: 0.00
            Seq primer:  40nl3 fwd, ET from Amersham
            High quality sequence stop: 496.
FEATURES              Location/Qualifiers
     source           1..544
                      /organism "Homo sapiens"
                      /db_xref "taxon:9606"
                      /clone "IMAGE:1525123"
                      /clone_lib "Soares_NFL_T_GBC_S1"
                      /lab_host "DH10B"
                      /note "Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site 1: Not I; Site 2: Eco RI;
                      Equal amounts of plasmid DNA from three normalized
                      libraries (fetal lung NbHLF9W, testis NHT, and B-cell
                      NCI_CGAP_GCB1) were mixed, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR amplified cDNAs from pools of 5,000 clones made
                      from the same 3 libraries. The pools consisted of
                      I.M.A.G.E. clones 297480-302887, 680242-687239,
                      726408-728711, and 729096-731899. Subtraction by Bento
                      Soares and M. Fatima Bonaldo. "
BASE COUNT       195 a    110 c    121 g    118 t
ORIGIN


alignment scores:
             Quality:  233.50             Length:      86
               Ratio:  3.538                Gaps:       1
 Percent Similarity:  76.744   Percent Identity: 50.000

alignment block:
 DS-09-126-945B-2 x AA913571   ..

 Align seq 1/1  to: AA913571  from: 1  to: 544

      246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 262
          |||:::   |||:::|||||:::|||||      ||||||
      117 GGAAACACAATTTATCTTTGGCAGTTTTTACTGGCACTGCTCCAGGACAA 166

      262 oHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 279
            :::      ::::::|||:::|||   ::::::|||||||||||||
      167 GGCTACTTCTCCTAAATACATCAAGTGGACCCAGCGAGAGAAAGGCATTT 216

      279 heLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLys 295
          |||||::   ||||||      |||:::|||||||||||   :::|||
      217 TTAAATTGGTGGATTCTAAAGCAGTGTCCAGTTGTGGGGGAAGCACAAA 266

      296 AsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTy 312
          |||:::|||   ||||||||||:::   ::::::|||:::::|||   ||
      267 AACAAACCTGATATGAATTATGAGACCATGGGAAGCAGCACTCAGGTACTA 316

      312 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVal1 329
          ||||:::::|||||||:::   |||   :::   :::|||||||||||||
      317 TTACCAAAGGGGTATTCTGGCAAAAGTGGAA...GGTCAGCGCTTGGTGT 363

      329 yrGlnPhe 331
          ||||||||
      364 ATCAGTTT 371

seq name: qb_est99:BG327400

seq documentation block:
LOCUS       BG327400      656 bp    mRNA            EST       27-FEB-2001
DEFINITION  602425874F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4563786 5',
            mRNA sequence.
ACCESSION   BG327400
VERSION     BG327400.1  GI:13113838
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   NIH-MGC Sequencing Group.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cmaps-r@mail.nih.gov
            Tissue Procurement: UCSD/ITP
            cDNA Library Preparation: Ling Hong/Robin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1278  row: e  column: 19
            High quality sequence stop: 654.
FEATURES              Location/Qualifiers
     source           1..656
                      /organism "Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone "IMAGE:4563786"
                      /clone_lib "NIH_MGC_14"
                      /tissue_type "renal cell adenocarcinoma"
                      /lab_host "DH10B (phage resistant)"
                      /note "Organ: kidney; Vector: pCMV7; Site 1: XhoI; Site 2:
                      EcoRI; cDNA made by oligo dT priming.  Directionally
                      cloned into EcoRI/XhoI sites using the following 5'
                      adaptor: GGCACGAG(G).  Size selected >500bp for average
                      insert size 1.8kb.  Library constructed by Ling Hong in
                      the laboratory of Gerald M. Rubin (University of
                      California, Berkeley) using  ZAP cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT       186 a    182 c    180 g    107 t      1 others
ORIGIN


alignment_scores:
             Quality:  233.50             Length:      86
               Ratio:  3.485                Gaps:       1
 Percent Similarity:  77.907   Percent Identity: 48.837

alignment_block:
 DS-09-126-945B-2 x BG327400   ..

 Align seq 1/1  to: BG327400  from: 1  to: 656

      246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 262
          |||:::   |||:::|||||||||||||||      ||||||
      252 GGCAGCACCATCIATCTGTGGGCAGTTCCTCCTGGCTCTTCTGCAAGACAG 301

      262 oHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 279
            ::::::      ::::::|||:::|||   ::::::|||||||||||||
      302 AAACACCTGTCCCAAGTACATCAAGTGGACCCAGCGAGAGAAAGGCATCT 351

      279 heLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLys 295
          |||||::   ||||||      |||:::|||||||||   :::|||
      352 TCAAACTGGTGGACTCCAAAGCTGTGTCCAAGCTGTGGGGGAAGCAGAAA 401

      296 AsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTy 312
          |||:::|||   ||||||||||:::   ::::::|||:::::|||   ||
      402 AACAAGCCTGACATGAACTATGAGACAATGGGGCGGCAGCTAAGATACTA 451

      312 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVal1 329
          ||||:::::|||||||:::   |||   :::   :::|||||||||||||
      452 CTACCAAAGAGGCATACTGGCCAAAGTGGAA...GGGCAGAGGCTGGTGT 498

      329 yrGlnPhe 331
          ||||||||
      499 ACCAGTTT 506

seq name: qb_est81:BE951696
```

```
seq_documentation_block:
LOCUS       BE951696      705 bp    mRNA            EST       04-OCT-2000
DEFINITION  UI-M-CC0-ayd-f-11-0-UI.s1 NIH_BMAP_Ret1 Mus musculus cDNA clone
            UI-M-CC0-ayd-f-11-0-UI 3', mRNA sequence.
ACCESSION   BE951696
VERSION     BE951696.1  GI:10591309
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791 806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9644, Bethesda, MD
            20892-9644, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            oligo-dT track not found, Not I site shown in terminal of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
            clones from RESEARCH GENETICS. It should be noted that Bento Soares
            is generating a small number of additional specialized
            non redundant arrays of BMAP cDNAs whose availability will be
            considered under appropriate and limited collaborative arrangements
            Seq primer: M13 Forward
            POLYA No.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism "Mus musculus"
                     /strain "C57BL/6J"
                     /db_xref "taxon:10090"
                     /clone "UI-M-CC0-ayd-f-11-0-UI"
                     /clone_lib "NIH_BMAP_Ret1"
                     /dev_stage "13.5 days pc"
                     /lab_host "DH10B (Life Technologies)"
                     /note "Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; the
                     NIH_BMAP_Ret1 library is 5' Stat from mouse eye cells, at
                     retina tissue. For a detailed description of the library
                     from which this clone was derived, please visit our web
                     site at brainest.enq.uiowa.edu.
                     TAG_SEQ None found"
BASE COUNT      216 a    171 c    199 q    119 t
ORIGIN


alignment_scores:
            Quality:  233.50             Length:     86
              Ratio:  3.485                Gaps:      1
 Percent Similarity:  77.907   Percent Identity: 48.837

alignment_block:
 US-09-126-945B-2 x BE951696

 Align seq 1/1  to: BE951696  from: 1  to: 705

     246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 282
         ||||::   |||:::|||||||::|||||      |||||
     257 GGCAGCACCATTTATCTGTGGGAGTTCCTCCTGGCCTCTCTGCAAGATAC 306

     282 eHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 279
         ::::::    ::::::|||:::|||   :::::|||||||||||||
     257 AAACACCTGCCCCAAGTACATTAAGTGGACGCAGAGAGAAGGGCATCT 306

     279 heLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLys 295
         |||||::   ||||||     |||::::::|||||||||   ::|||
     307 TCAAGTTGGTGGACTCCAAGGCTGTGTCCAAGCTGTGGGGCAAGCAGAAA 356

     296 AsnArgPheAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTy 342
         |||::::||   ||||||||||:::   ::::::|||||||||||||   ||
     357 AACAAACCTGACATGAACTACGAGAAG ATGGGCCAGCCCTCAGATATTA 406

     312 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVal1 329
         ||||::::::|||||||:   |||   :::   :::|||||||||||||
     407 TTATCAAAGAGGAATCCTGGCCAAAGTGGAA...GGGCAGAGGCTGGTAT 453

     329 yrGlnPhe 331
         ||||||||
     454 ACCAGTTC 461

seq_name: gb_est53:AW949623

seq_documentation_block:
LOCUS       AW949623      556 bp    mRNA            EST       01-JUN-2000
DEFINITION  EST361693 MAGE resequences, MAGA Homo Sapiens cDNA, mRNA sequence.
ACCESSION   AW949623
VERSION     AW949623.1  GI:8139255
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 4
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib="MAGE resequences, MAGA"
                     /note "Vector: pBluescriptSKm"
BASE COUNT      201 a     95 c    129 q    131 t
ORIGIN


alignment_scores:
            Quality:  231.50             Length:     86
              Ratio:  3.508                Gaps:      1
 Percent Similarity:  76.744   Percent Identity: 48.837

alignment_block:
 US-09-126-945B-2 x AW949623  ..

 Align seq 1/1  to: AW949623  from: 1  to: 556

     246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 282
         ||||::   :::|||||||:::|||||   :::||||||
      64 GGAAACACAACCTATTTGTGGGAGTTCCTCCTAGATCTACTCCTCAAGATAA 113

     282 eHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 279
         ::::::      |||:::|||:::|||   :::::|||||||||||||
     114 AAATACTTGTCCCAGCATATTAAATGGACTCAGAGAGAAAAAGGCATAT 163

     279 heLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLys 295
         |||||::   ||||||     |||   :::|||||||||   ::|||
     164 TCAAGCTGGTGGATTCAAAGGCTGTCTATAAGCTTTGGGGAAAGCAIAAG 213
```

```
      296 AsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTy 312
          |||:::|||   ||||||||||:::   :::::|||:::::::|||   ||
      214 AACAAACCAGACATGAACTATGAAACCATGGGACGAGCTTTGAGATACTA 263

      312 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValT 329
          ||||::::::|||||||::   |||   :::   :::|||||||||||||
      264 CTACCAAAGGGCAATTCTTGCAAAGGTTGAA...GGACAGAGGCTTGTAT 310

      329 yrGlnPhe 331
          ||||||||
      311 ATCAGTTC 318

seq_name: qb_est27:AI958988

seq_documentation_block:
LOCUS        AI958988       562 bp    mRNA            EST       20-AUG-1999
DEFINITION   fd06b10.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
             qb:MB2882 ETS-RELATED TRANSCRIPTION FACTOR ELF 1 (HUMAN);, mRNA
             sequence.
ACCESSION    AI958988
VERSION      AI958988.1  GI:5751701
KEYWORDS     EST.
SOURCE       zebrafish.
  ORGANISM   Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE    1  (bases 1 to 562)
  AUTHORS    Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
             ,S., Hillier,L., Kucaba,T., Martin,J., Bock,C., Wylie,T., Underwood
             ,K., Stoptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
             Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,E., Ritter,E.,
             Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
             and Wilson,R.
  TITLE      WashU Zebrafish EST Project 1998
  JOURNAL    Unpublished (1998)
COMMENT      Other ESTs: fd06b10.x1
             Contact: Stephen L. Johnson
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: zbrafish@watson.wustl.edu
             cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
             Matthew Clark  DNA Sequencing by: Washington University Genome
             Sequencing Center Clone distribution: Genome Systems, St. Louis,
             Missouri (web address: www.genomesystems.com) (email contact:
             info@genomesystems.com) and Research Genetics, Huntsville, Alabama
             (web address: www.resgen.com) (email contact: info@resgen.com) and
             ResSourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
             www.rzpd.de)
             Seq primer: T3 ET from Amersham
             High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1..562
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF'"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I   oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT3']
                     double stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT       177 a    125 c    142 g    117 t      1 others
ORIGIN


alignment_scores:
            Quality:  231.50              Length:     82
              Ratio:    3.508               Gaps:      1
 Percent Similarity:   80.488   Percent Identity: 50.000

alignment_block:
 03 09 126 945B 2 x AI958988  ..

 Align seq 1/1  to: AI958988  from: 1  to: 562

      250 HisLeuTrpAlaPheLeuLysGluLeuLeuLysProHisSerTyrGl 266
          :::||||||::|||||   :::|||||          ::::::
      114 TACCTGTGGGAGTTCCTGTTGGACCTGCTGCAGGATAAGAACACCTGTCC 163

      266 yArgPheIleArgProLeuAsnLysGluLysGlyIlePheLysIleGlnA 283
          |||:::|||::|||   ::::::|||||||||||||||||:::   |
      164 CAGATACATCAAGTGGACACAGAGAGAAGGGCATCTTTAAACCTGGTGG 213

      283 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299
          |||||      |||:::::|||||||||   :::|||||||:::|||
      214 ACTCCAAACCCCTCCAAAACTCTGCGCCAAACACAAAAACAAACCACAC 263

      300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
          ||||||||||:   :::::|||:::::|||:::|||||||||:::::||
      264 ATGAACTATGAAACCATGGGTAGACGCTCAGGGATTACTATCAGAGAGG 313

      316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          |||:::   |||   :::   :::|||||||||||||||||||||
      314 CATCTTGGCTAAAGTCGAG...GGTCAGAGGTTGGTGTATCAGTTT 356

seq_name: qb_est4:AA261331

seq_documentation_block:
LOCUS        AA261331       555 bp    mRNA            EST       18-MAR-1997
DEFINITION   va35d02.r1 Soares mouse 3NME12.5 Mus musculus cDNA clone
             IMAGE:733847 5' similar to qb:M25269 PROTEIN ELK-1 (HUMAN);
             qb:X87257 M.musculus mRNA for elk1 protein (MOUSE);, mRNA sequence.
ACCESSION    AA261331
VERSION      AA261331.1  GI:1895875
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 555)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Gojsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Stoptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
  TITLE      The WashU-HHMI Mouse EST Project
  JOURNAL    Unpublished (1996)
COMMENT      Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:450895
```

```
            Seq primer: -28ml3 rev2 ET from Amersham
            High quality sequence stop: 431.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism "Mus musculus"
                     /strain "[illegible]"
                     /sub_strain "[illegible]"
                     /clone="IMAGE:743347"
                     /clone_lib "Soares mouse 3NMEl2 5"
                     /sex "Unknown"
                     /tissue_type="fetus"
                     /dev_stage "12.5dpc total fetus"
                     /lab_host="DH10B"
                     /note "Organ: whole fetus; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker, Site_1: Not I, Site_2: Eco RI.
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCTTTTTTTTTTTTTT
                     3'], on total mouse RNA [provided by Mitzen F. Wayne
                     State Univ.]; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       121 a    154 c    168 g    112 t
ORIGIN


alignment_scores:
            Quality:  230.00              Length:     414
              Ratio:  2.771                 Gaps:       5
 Percent Similarity:  72.807   Percent Identity: 45.614

alignment_block:
 US-09-126-945B-2 x AA261331   ..

 Align seq 1/1  to: AA261331  from: 1  to: 555

      221 ProGlyAlaIleHis...TyrCysAlaSerThrSerGluGluSerTrpTh 236
          |||   :::   |||   :::       ::|||   :::   ||||||
       57 CCGCCGGAGCTGCCACTGCCACCGCCGCCGCCACTCGCAGGTACTCCCTGG.. 104

      236 rAspSerGluValAspSerSerCysSerGlyGlnPro...IleHisLeuT 252
            :::   |||:::::::||         :::|||   :::   ||||
      105 ....AATGGCGTGAGTGCTTCCCTAGTGATGGACCCATCTGTGACGCTGT 150

      252 rpGlnPheLeuLysGlnLeuLeuLeuLysProHisSerTyrGlyArgPhe 268
          ||||||||||||   :::   |||||:::   ::::::   |||::::::
      151 GGCAGTTTCTGCTGCAG...CTTCTGAGAGAACAAGGTAATGGCCACATC 197

      269 IleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAl 285
          |||   |.|   :::::::::   |||   |||||:::   ||:::
      198 ATCTCCTGGACCTCACGGGATGGTGGTCAGTTCAAGTTGGTGGATGCAGA 247

      285 aGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsn 302
          :::||||||||||||||||||:::||||||||:::      |||||||
      248 GGAGCTCGCCCCCCTATGCCACCCCCAACAACAACACCAACATCAATT 297

      302 yrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIle[illegible] 318
          ||||||| ||||||||||:::::|||   ||||||   |||:::|||||
      298 ACGACAAGCTTAGCCGGGCCTTGCGGTACTACTATGATAAGAATATCATC 347

      319 ArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal 332
          ||:::   :::   :::|||:::::|||||||||||||||
      348 CGCCAAGTTCAGC...GGGCAGAAGTTTGTCTACCAGTTTCTC 386

seq_name: qb_est100:BG393711

seq_documentation_block:
LOCUS       BG393711      788 bp    mRNA            EST       12-MAR-2001
DEFINITION  602412140F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4541048 5',
            mRNA sequence.
ACCESSION   BG393711
VERSION     BG393711.1  GI:13287159
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   [illegible]
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10471  row: p  column: 09
            High quality sequence stop: 756.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:4541048"
                     /clone_lib "NIH_MGC_92"
                     /tissue_type "embryonal carcinoma, cell line"
                     /lab_host "DH10B (phage-resistant)"
                     /note "Organ: testis; Vector: pCMV SPORT6; Site_1: Not I;
                     Site_2: Sall; Cloned unidirectionally; oligo dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT       156 a    256 c    242 g    134 t
ORIGIN


alignment_scores:
            Quality:  249.00              Length:     150
              Ratio:  2.500                 Gaps:       8
 Percent Similarity:  61.333   Percent Identity: 39.333

alignment_block:
 US-09-126-945B-2 x BG393711   ..

 Align seq 1/1  to: BG393711  from: 1  to: 788

      192 ArgGlnArgSerPro...LeuGlyGlyAspValLeuHisAlaHisLeuAs 207
          |||   |||:::|||   |||||||||
        6 CGCGCTGGGAACCCAGGCCTGGGGGGCGGT..................... 35

      207 pIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAlaI 224
                ::|||||||:::              :::|||   :::
       36 .........GGGGCCGCGTAT.....GGAGCCCCCGCCCCCCGGAGCT 69

      224 leHisTyrCysAla.............SerThrSerGlyGluSerTrp... 235
          :::::::|||                ::|||   :::      |||
       70 GCCAACATCGCAACGCCACGCCACGCTACACACAGGTACCCCTGGGAT 119

      236 .........IhrAspSerGluValAspSerSerCysSerGlyGlnProIl 249
                  |||           ::|||   |||              ::
      120 GGCGTGAGCACTCCCCCAGCGATGGACCCATCT.............GTC 154

      249 eHisLeuTrpGlnPheLeuLysGlnLeuLeuLeuLysProHisSerTyrG 266
          :   |||||||||||||||   :::   |||||:::   ::::::   |
      155 GACGCTGTGGCAGTTTCTGCTGCAG...CTGCTGAGAGAGCAAGGCAATG 201

      266 lyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGlu 282
          |||::::|||   |||  :::::::::   |||   |||||:::
      202 GCCACATCATCCTGGACCTCACGGGATGGTGGTCAATTCAAGTTGGTGG 251

      283 AspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAl 299
          ||:::   :::|||||||||||||||:::|||||||||:::
```

```
    252 GATGCAGAGGAGGTGGCCCGGCTGTGGGGGCTACGCAAGAACAAGACCAA 301

    299 aMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysG 316
         |||||||||||||||||||||||:::::|||    ||||||    |||:
    302 CATGAATTACGACAAGCTCAGCCGGGCCTTGCGGTACTACTATGACAAGA 351

    316 lyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal 332
         ::||||||||||||    :::     :::|||::::::||||||:::||||||
    352 ACATCATCCGCAAGGTGAGC...GGCCAGAAGTTCGTCTACAAGTTTGTG 398

seq_name: gb est29:AL560620

seq documentation block:
LOCUS       AL560620       898 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL560620 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL003YJ09 5
            prime, mRNA sequence.
ACCESSION   AL560620
VERSION     AL560620.1  GI:12907259
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DL003YJ09"
                     /clone_lib="LTI_NFL010_BC2"
                     /sex="male"
                     /tissue_type="B cells from Burkitt lymphoma"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI oligo(dT) primer. Five prime end
                     enriched, double stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      181 a    292 c    274 g    151 t
ORIGIN


alignment scores:
           Quality:   228.50          Length:     121
             Ratio:   2.821             Gaps:       5
 Percent Similarity:  66.942  Percent Identity:  42.149

alignment block:
 US-09-126-945B-2 x AL560620 ..

 Align seq 1/1  to: AL560620  from: 1  to: 898

    220 SerProGlyAlaIleHisTyrCysAla............SerThrSerGl 232
         :::|||   :::   ::::::|||            :::|||    ::
     53 GCCCCCGGAGCTGCCAACATTGCCAACGCCACCGCCACGCTACACAG 102

    232 nGluSerTrp...........ThrAspSerGluValAspSerSerCysS 245
         :       |||              |||        :::|||   |||
    103 GTACCCCTGGGATGGCGTGAGCACTCCCCAGCGATGGACCCATCT.... 148

    245 erGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLys 261
                 :::   |||||||||||||||   :::    ||||||:::

    149 ...........GTGACGCTGTGGCAGTTTCTGCTGCAG...CTGCTGAGA 184

    262 ProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGlnLysGlyIl 278
            ::::::    |||:::::::|||    |||    :::::::::    |||
    185 GAGCAAGGCAATGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGA 234

    278 ePheLysIleGlnAspSerAlaGlnValAlaArgLeuTrpGlyIleArgL 295
         ||||||:::   |||:::    :::||||||||||||||||||::::||||
    235 ATTCAAGCTGGTGGATGCAGAGGAGGTGGCCCGGCTGTGGGGGCTACGCA 284

    295 ysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGln 311
         ||||:::        ||||||||||||||||||||||||::::::|||
    285 AGAACAAGACCAACATGAATTACGACAAGCTCAGCCGGGCCTTGCGGTAC 334

    312 TyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVa 328
         ||||||    |||:::||||||||||||     :::      :::|||::::::||
    335 TACTATGACAAGAACATCATCCGCAAGGTGAGC...GGCCAGAAGTTCGT 381

    328 lTyrGlnPheVal 332
         |||:::||||||
    382 CTACAAGTTTGTG 394

seq_name: gb est88:BF450651

seq_documentation block:
LOCUS       BF450651       642 bp    mRNA            EST       29-DEC-2000
DEFINITION  uz68g07.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3674268 5'
            similar to TR:Q15724 Q15724 NERF-1B. [1] ;, mRNA sequence.
ACCESSION   BF450651
VERSION     BF450651.1  GI:11516820
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html/resources.shtml

            MGI:1435036
            Seq primer: -40RP from Gibco
            High quality sequence stop: 437.
FEATURES             Location/Qualifiers
     source          1..642
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3674268"
                     /clone_lib="NCI_CGAP_Mam6"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      212 a    123 c    145 g    162 t
ORIGIN


alignment_scores:
```

```
            Quality:  228.00              Length:     224
              Ratio:  2.018                 Gaps:       6
 Percent Similarity:  50.416   Percent Identity: 27.679

alignment_block:
 US-09-126-945B-2 x BF450651   ..

 Align seq 1/1  to: BF450651  from: 1  to: 642

      108 ValProGlyGlyLeuThrLeuGlnGluHisSerLeuGlnGlnValGluSe 124
          ||||||      ::::::      |||                 ::::::
      103 GTCCCTCTTGTATATCAACTCCAGAGTTTATCCATGCTGCGATGAGCCC 152

      124 rMetValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuL 141
            ||:::   |||::::::   ::::::::::|||
      153 GGATGTCATCACAGAGACTGTAGTGGAGGTGTCCACAGAA.......... 192

      141 euAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGluLys 157
                :::::|||||||||   ||||||
      193 ........GAGTCTGAGCCGATGGATGCCTCGCCT.............. 219

      158 IrpLeuLeuIrpThrGluHisGluTyrArqLeuProProMetGlyLysAl 174
                                      ::|||       ::
      220 ............................ATCCCTACTTCACCAGACAG 239

      174 aPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnP 191
          :   :::   :::      |||:::         ::::::::::::
      240 GCAACAA?AAATGAAAAACAAAAAA          GTCCCCGTAAA???AA 282

      192 heArqGluArqSerProLeuGlyGlyAspValLeuHisAlaHisLeuAsp 207
            |||:::||||||:::::::::            ::::||
      281 AGA???AGCAGTCACCA?TTCCAATGGG......TCTCCTGAGTTA?GA 324

      208 IleTrpLysSerAlaAlaTrpMetLysGluArqThrSerProGlyAlaIl 224
          |||   |||::::||      |||
      325 ATCAAGAAGAAAGCCAGAGAAGGAAAA..................... 351

      224 eHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluValA 241

      351 ................................................. 351

      241 spSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGlu 257
                    ||:::      :::|||||||:|||||   :::
      352 ............GGAAACACAACCTATCTGTGGGAGTTTCTTTAGAT 387

      258 LeuLeuLeuLysProHisSerTyrGlyArgPheIleArrTrpLeuAsnLy 274
          ||||||      ::::::      ||:::|||:::|:|   :::::
      388 TTGCTTCAAGATAAAAATACTTGTCCCAGGTATATTAAATGGACTCAAAG 437

      274 sGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArqLeuI 291
          :|||||||||||||||||||:::   ||||||     |||:::::::|||
      438 AGAAAAGGGCATATTCAAACTTGTGGACTCAAAGGCTGTCTCGAAGCTCT 487

      291 rpGlyIleArqLysAsnArqProAlaMetAsnTyrAspLysLeuSerArq 307
          |||||   :::|||||:::::|   ||||||||:::   :::::|||
      488 GGGGCAAGCATAAGAACAAACCAGATATGAACTATGAAACCATGGGGCGA 537

      308 SerIleArqGlnTyrTyrLysLysGlyIleIleArqLysPr Arg|:|G  ?1
          :::::|||   ||||||:::::|||||||:::   |||   :::
      538 GCGTTAAGATATTACTACCAAAGGGGAATTCTTGCAAAGGTTGAA...TG 584

      324 rGlnArqLeuValTyrGlnPhe 337
            |||   |||||||||||||||
      585 ACAGATGCTTGTATATCAGTTC 606

seq_name: qb_est80:BE888101

seq_documentation_block:
LOCUS       BE888101       589 bp    mRNA            EST       20-OCT-2000
DEFINITION  601511610F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3912915 5',
            mRNA sequence.
ACCESSION   BE888101
VERSION     BE888101.1  GI:10344066
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 589)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9732  row: d  column: 04
            High quality sequence stop: 588.
FEATURES             Location/Qualifiers
     source          1..589
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3912915"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV SPORT6; Site_1: Not1;
                     Site_2: Sal1; Cloned unidirectionally.  Primer: oligo dT;
                     Average insert size 2.1 kb.  "
BASE COUNT       125 a    180 c    183 g    101 t
ORIGIN


alignment_scores:
            Quality:  227.50              Length:     120
              Ratio:  2.844                 Gaps:       5
 Percent Similarity:  66.667   Percent Identity: 42.500

alignment_block:
 US-09-126-945B-2 x BE888101   ..

 Align seq 1/1  to: BE888101  from: 1  to: 589

      221 PrGlyAlaIleHisTyrCysAla             SerThrSerGluGl 243
          |||  :::   ::::::|||                 :::|:|   :::
       84 CCCCGGAGCTGCAACATGCCAACGCCACCGCCACGCTACACACAGGTA 83

      233 uSerTrp.............ThrAspSerGluValAspSerSerCysSerG 246
            |||               |||           ::::||   |||
       84 CCCCTGGGATGGCGTGAGCACTCCCCCAGCGATGGACCCATCT....... 126

      246 lyGlnProIleHisLeuIrpGlnPheLeuLysGluLeuLeuLeuLysPro 262
                :::   |||||||||||||||   :::   ||||||::
      127 .......GTGACGCTGTGGCAGTTTCTGCTGCAG...CTGCTGAGAGAG 165

      263 HisSerTyrGlyArgPheIleArrTrpLeuAsnLysGluGlyGlyIlePh 279
          ::::::  |||:::::::||   |||   :::::::::  |||   ||
      166 CAAGGCAATGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGAATT 215

      279 eLysIleGluAspSerAlaGlnValAlaArqLeuTrpGlyIleArqLysA 296
          ||||:::   ||||::   :::|||||||||||||||||||:::|||||
      216 CAAGCTGGTGGATGCAGAGGAGGTGGCCCGGCTGTGGGGGCTACGCAAGA 265

      296 snArqProAlaMetAsnTyrAspLysLeuSerArqSerIleArqGlnTyr 312
          ||:::      |||||||||||||||||||||||:::::|||   |||
      266 ACAAGACCAACATGAATTACGACAAGCTCAGCCGGGCCTTGCGGTACTAC 315

      313 TyrLysLysGlyIleIleArqLysProAspIleSerGlnArqGlnValTy 329
          |||   |||:::|||||||||||   :::   :::|||:::::|||||
      316 TATGACAAGAACATCATCCGCAAGGTGAGC...GGCCAGAAGTTCGTCTA 362
```

```
      329 rGlnPheVal 342
          |:::|||||
      363 CAAGTTTGTG 372

Seq_name: qb_est4:AA228986

Seq_documentation_block:
LOCUS        AA228986       441 bp    mRNA            EST       21 AUG-1997
DEFINITION  ne50f12.s1 NCI_CGAP_Pr4 Homo sapiens cDNA clone IMAGE:1011599
            similar to qb:M82882 ETS-RELATED TRANSCRIPTION FACTOR ELF-1 (HUMAN
            );; mRNA sequence.
ACCESSION    AA228986
VERSION      AA228986.1  GI:1852019
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 441)
  AUTHORS    NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
             , Michael Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www.bio.llnl.gov/bbrp/image/image.html
             Seq primer:  41mi3 fwd. ET from Amersham
             High quality sequence stop: 249.
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:1011599"
                     /clone_lib "NCI_CGAP_Pr4"
                     /sex "Male"
                     /dev_stage "45 years old"
                     /lab_host "DH10B"
                     /note "Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse treated, total cellular RNA obtained from 5,000 10
                     ,000 microdissected cells histologically-determined to be
                     fully malignant prostate cancer cells.  Double-stranded
                     cDNA was ligated to EcoRI adaptors, 5 cycles of PCR
                     applied to the cDNA with an adaptor-specific primer, and
                     the resulting PCR product subcloned into pAMP10 by the
                     UDG cloning method (Life Technologies).  Average insert
                     size is 900 bp.  NOTE: Not dir. bleed/, cbleed.  This
                     library was constructed by  David Krizman."
BASE COUNT       144 a      80 c     110 g     107 t
ORIGIN


alignment scores:
            Quality:  226.50            Length:      86
              Ratio:    3.485             Gaps:       1
 Percent Similarity:  75.581   Percent Identity: 48.837

alignment block:
 US-09-126-945B-2 x AA228986   ..

 Align seq 1/1  to: AA228986  from: 1  to: 441

      246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 262
          |||:::  |||:::  |||:::||||||     ||||||
       22 GGAAACACAATTTATTCTTGGGAGTTTTACTGGCACTGCTCCAGGACAA 71

      262 aHisProTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 229
          :::   ::::::|||:::|||  :::::|||||||||||||
       72 GGCTACTTGTCCTAAATACATCAAGTGGAGCCAGCGAGAGAAAGGCATTT 121

      229 heLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLys 295
          ||||:::  ||||||     |||:::|||||||||||  :::|||
      122 TTAAATTGCTGGATTCTAAAGCAGTGTCCAGGTTGTGGGGGAAGCACAAA 171

      296 AsnArgProAlaMetAspTyrAspLysLeuSerArgSerIleArgGlyTy 342
          |||:::|||  ||||||||||::  ::::::|||::::::|||  ||
      172 AACAAACCTGATATGAATTATGAGACCATGGGAACAGCACTCAGGIACTA 221

      312 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValT 329
          |||:::::|||||:::  |||  :::  :::|||||||||||||
      222 TTACCAAAGGGGIATTCTGGCAAAAGTGGAA...GGTCAGCGCTTGGTGT 268

      329 yrGlnPhe 331
          ||||||||
      269 ATCAGTTT 276

Seq_name: qb_est87:BF379908

Seq_documentation_block:
LOCUS        BF379908       301 bp    mRNA            EST       27-NOV-2000
DEFINITION  CM4-UT0043-260900-561-G05 UT0043 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BF379908
VERSION      BF379908.1  GI:11869033
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 301)
  AUTHORS    Dias Neto,E., Correia Correa,P., Verjovski-Almeida,S., Briones,M.R.,
             Naqui,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE    20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/getEtm12.pl?t1 CM1&t2 CM4 UT0043
             260900-561-G05&t3 2000-09-26&t4 1)
             Seq primer: puc 18 forward
             High quality sequence start: 2
             High quality sequence stop: 300.
FEATURES             Location/Qualifiers
     source          1. .301
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone_lib "UT0043"
                     /dev_stage "Adult"
                     /note "Organ: uterus tumor; Vector: pucl8; Site_1: SmaI;
                     Site_2: SmaI, A mini library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT        51 a      82 c      68 g     100 t
ORIGIN
```

```
alignment_scores:
            Quality:  225.50              Length:     86
              Ratio:  1.417                 Gaps:      1
 Percent Similarity:  76.744   Percent Identity: 47.674

alignment_block:
 US-09-126-945B-2 x BF379908/rev   ..

 Align seq 1/1  to reverse of: BF379908  from: 1  to: 301

      246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 262
          |||:::   |||:::|||||||:::|||||       ||||||
      278 GGCAGCACCATCTATCTGTGGGAGTTCCTCCTGGCTCTTCTGCAACACAC 224

      262 eHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 279
            ::::::      ::::::|||:::|||   :::::|||||||||||||
      223 AAACACCTGTCCCAAGTACATCAAGTGGACCCAGCGAGAGAAAGGCATCT 174

      279 heLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLys 295
          |||||::   ||||||      |||:::::|||||||   :::|||
      173 TCAAACTGGTGGACTCCAAAGCTGTGTCCAAGCTGTGGGGGAAGCAGAAA 124

      296 AsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTy 312
          |||:::|||:::||||||||||:   :::::|||:::::|||   ||
      123 AACAAGCCTGGCATGAACTATGACAAATGGGGTGGCACTAAGATACTA 74

      312 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVal 329
          ||||:::::|||||      |||          :::|||||:::||||
       73 CTACCAAAGAGGCATACCGGGCAAAGTCAAA...GGCCAGAGGATGGTGT 27

      329 yrGlnPhe 331
          |||||||
       26 ACCAGTTT 19

seq_name: qb_est84:BF180927

seq_documentation_block:
LOCUS       BF180927      961 bp    mRNA            EST       31-OCT-2000
DEFINITION  601807167F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4037776 5',
            mRNA sequence.
ACCESSION   BF180927
VERSION     BF180927.1  GI:11059069
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 961)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9315  row: f  column: 17
            High quality sequence stop: 654.
FEATURES             Location/Qualifiers
     source          1. .961
                     /organism="Mus musculus"
                     /strain "C57/B6"
                     /db_xref "taxon:10090"
                     /clone="IMAGE:4037776"
                     /clone_lib="NCI_CGAP_Mam5"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host "DH10B"
                     /note "Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sal1;
                     Site_2: Not1; Cloned unidirectionally.  Primer: oligo dT.
                     Library constructed by Life Technologies.  Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT      282 a    252 c    245 g    182 t
ORIGIN


alignment_scores:
            Quality:  225.50              Length:    175
              Ratio:  1.895                 Gaps:      5
 Percent Similarity:  68.000   Percent Identity: 38.714

alignment_block:
 US-09-126-945B-2 x BF180927   ..

 Align seq 1/1  to: BF180927  from: 1  to: 961

      165 GlnTyrArgLeuProProMetGlyLysAlaPhe......GlnGluLeuAl 179
          ||||||:::|||              :::|||           ::::::::
       17 CAGTACAAGCTTGATGCCAACTGCATCTCCTTCTGTCACTTCAACATCAG 66

      179 aGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerP 196
          :|||   :::|||||:::||||:::::|||||||   :::   :::
       67 CGGCCTGCAGCTCTGCAGCATGACGCAGCAGGACTTCATTGAGGCAGCCG 116

      196 rgLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerAla 212
            :::   |||:::   |||:::      |||:::      ::::||
      117 GCATCTGTGGGGAGTAGCTGTACTTCATTCTCCAGAACATTCGCTCGCAA 166

      213 AlaTrpMet.............lysGluArgThrSerProGlyAlaIleHi 225
          ::::::                |||:::|||:::   |||
      167 GGTTACTCCTTTTCAATGATGCTGAAGAAACCAAGACTGGCATCAAAGA 216

      225 rTyrCysAlaSerThrSer...GlnGlnSerTrp...ThrAspSerGluVa 240
          |||:::   |||:::     |||:::::|||        :::|||:::::
      217 CTATGCTGATTCCACTTGCTTGAAAACAACTGGCATCAAGACTCAACACT 266

      240 lAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysG 257
          :::::::      |||   |||      ||||||||||:::   :::::::
      267 GTCACAGCGAACAAGCCTCCAAAGTTCTCACCTGTGGGAATT.GTCAGAG 315

      257 luLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsn 273
          ::||||||||:::|||:::::          :::::::::||   :::
      316 ACTTGCTGCTGTCCCCTGAAGAGAACTGTGGCATCCTGGAATGGGAAGAC 365

      274 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLe 290
          :::|||:::||||||||||:::::      |||     :::|||:::::
      366 AGGAAGAGGGCATTTTCCGAGTGCTTAAGTCAGAAGCCCTGGCAAAGAT 415

      290 uTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerA 307
          :|||||||   |||   ::::::      |||:::|||:::||||||||||
      416 GTGGGCACAAAGGACGAACAATGACAGGATGACGTACGACAAGCTGAGCC 465

      307 rgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIle 323
          ||:::::::|||   ||||||||   |||||||||:::::   |||
      466 GAGCCCTGAGATACTACTATAAAAAGGGAATTCTGGAGCGGGTTGAC... 512

      324 SerGlnArgLeuValTyrGlnPhe 331
          :::|||||||||||:::|||
      513 ...CGGAGGTTAGTGTACAAATTT 533

seq_name: qb_est94:BF953438

seq_documentation_block:
LOCUS       BF953438      512 bp    mRNA            EST       22-JAN-2001
DEFINITION  RC3-NN0195-121100-022-g10 NN0195 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF953438
VERSION     BF953438.1  GI:12370713
KEYWORDS    EST.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brumstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun Sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11 2704922
            Fax: +55 11 2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/qethtml2.pl?LI_RC4&L2_RC3_NN0195-
            121109-022-q10&t3_2000_11-12&t4_1)
            Seq primer: puc 18 forward
            High quality sequence start: 22
            High quality sequence stop: 512.
FEATURES              Location/Qualifiers
     source           1..512
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="NN0195"
                      /dev_stage="Adult"
                      /note="Organ: nervous normal; Vector: pucl8; Site 1: SmaI;
                      Site 2: SmaI; A mini library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716    Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT       110 a    156 c    141 g    104 t      1 others
ORIGIN


alignment scores:
            Quality:  224.00            Length:     84
              Ratio:  3.343               Gaps:      2
 Percent Similarity:  79.762   Percent Identity:  52.381

alignment block:
 US-09-126-945B-2 x BF953438   ..

 Align seq 1/1  to: BF953438  from: 1  to: 512

      249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
          :::   ||||||||||||||||   :::   ||||||:::   ::::::
      209 GTGACGCTGTGGCAGTTTCTGCTGCAG...CTGCTGAGAGAGCAACGCAA 259

      265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleG 282
           |||::::::|||   |||   :::::::::   |||   ||||||||::
      256 TGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGAATTCAAGCTGG 305

      282 IuAspSerAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgPro 298
          |||:::   :::|||||||||||||||||||:::||||||||||::::
      306 TGGATGCAGAGGAGGTGGCCCGGCTGTGGGGGCTACGCAAGAACAAGACC 355

      299 AlaMetAsnTyrAspLysLeuSerArgSerIleArgIleTyrTyrLysGly 315
          ||||||||||||||||||||||||||:::::|||   ||||||   ||
      356 AACATGAATTACGACAAGCTCAGCCGGGCCTTGCGGATACTACTATGACAA 405

      315 sGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGluPheV 332
```

```
          :::|||||||||||||   :::   :::|||:::::::||||||:::||||
      406 GAACATCATCCGCAAGCTGAG...GGCCAGAAGTTCCTCTACAAGTTTG 452

      332 al 332
          ||
      453 TG 454

seq name: qb_est31:AU129451

seq documentation block:
LOCUS       AU129451       674 bp    mRNA             EST       24-OCT-2000
DEFINITION  AU129451 NT2RP2 Homo sapiens cDNA clone NT2RP2005559 5', mRNA
            sequence.
ACCESSION   AU129451
VERSION     AU129451.1  GI:10989805
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 674)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     RIKEN human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-4951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            RIKEN human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute, cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES              Location/Qualifiers
     source           1..674
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="NT2RP2005559"
                      /clone_lib="NT2RP2"
                      /cell_type="teratocarcinoma"
                      /cell_line="NT2"
                      /note="Vector: pMF18SFL3; mRNA from NT2 neuronal precursor
                      cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT       138 a    219 c    181 g    134 t      2 others
ORIGIN


alignment scores:
            Quality:  224.00            Length:     84
              Ratio:  3.343               Gaps:      2
 Percent Similarity:  79.762   Percent Identity:  52.381

alignment block:
 US-09-126-945B-2 x AU129451   ..

 Align seq 1/1  to: AU129451  from: 1  to: 674

      249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
          :::   ||||||:|||||||   :::   ||||||:::   ::::::
      121 GTGACGCTGTGGCAGTTCCTGCTGCAG...CTGCTGAGAGAGCAACGCAA 167

      265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleG 282
           |||::::::|||   |||   :::::::::   |||   ||||||||::
      168 TGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGAATTCAAGCTGG 217

      282 IuAspSerAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgPro 298
          |||:::   :::|||||||||||||||||||:::||||||||||::::
      218 TGGATGCAGAGGAGGTGGCCCGGCTGTGGGGGCTACGCAAGAACAAGACC 267
```

```
      299 AlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLy 315
          ||||||||||||||||||||||||||||||||||   ||||||   ||
      268 AACATGAATTACGACAAGCTCAGCCGGGCCTTGCGGTACTACTATGACAA 317

      315 sGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheV 332
          |:::|||||||||||   :::   :::|||:::::|:|||||:::|||
      318 GAACATCATCCGCAAGGTGAGC...GGCCAGAACTTCGTCTACAAGTTTG 363

      332 aI 332
          ||
      365 IG 366

Seq_name: qb_ost29:AL556333

seq_documentation_block:
LOCUS       AL556333      848 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL556333 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK001YM12 5
            prime, mRNA sequence.
ACCESSION   AL556333
VERSION     AL556333.1  GI:12898915
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 848)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqret@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..848
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DK001YM12"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies, Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8471
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      170 a    286 c    236 g    155 t      1 others
ORIGIN


alignment_scores:
            Quality:  224.00             Length:     84
              Ratio:  3.343                Gaps:      2
 Percent Similarity:  79.762   Percent Identity: 52.481

alignment_block:
 US-09-126-945B-2 x AL556333   ..

 Align seq 1/1  to: AL556333  from: 1  to: 848

      249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
          :::   |||||||||||||||   :::   |||||||::   ::::::
      249 GTGACGCTGTGGCAGTTTCTGCTGCAG...CTGCTGAGAGAGCAAGGCAA 295

      265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleG 282
          |||::::::|||   |||   :::::::::   |||   ||||||:::
      296 TGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGAATTCAAGCTGG 345

      282 luAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgPro 298
          |||:::   :::|||||||||||||||||||:::|||||:|||:::
      346 TGGATGCAGAGGAGGTGGCCTGGCTGTGGGGACTACGCAAGAACAAGACC 395

      299 AlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLy 315
          |||||||||||||||||||||||||||||||||   ||||||   ||
      396 AACATGAATTACGACAAGCTCAGCCGGGCCTTGCGGTACTACTATGACAA 445

      315 sGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheV 332
          |:::|||||||||||   :::   :::|||:::::|:|||||:::|||
      446 GAACATCATCCGCAAGGTGAGC...GGCCAGAACTTCGTCTACAAGTTTG 492

      332 aI 332
          ||
      493 IG 494

Seq_name: qb_ost29:AL555439

seq_documentation_block:
LOCUS       AL555439     1003 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL555439 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK008YM09 5
            prime, mRNA sequence.
ACCESSION   AL555439
VERSION     AL555439.1  GI:12897174
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1003)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqret@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..1003
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DK008YM09"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies, Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8471
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      209 a    313 c    311 g    168 t      2 others
ORIGIN


alignment_scores:
            Quality:  224.00             Length:     84
              Ratio:  3.343                Gaps:      2
 Percent Similarity:  79.762   Percent Identity: 52.481

alignment_block:
 US-09-126-945B-2 x AL555439   ..

 Align seq 1/1  to: AL555439  from: 1  to: 1003

      249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
          :::   |||||||||||||||   :::   |||||||::   ::::::
      142 GTGACGCTGTGGCAGTTTCTGCTGCAG...CTGCTGAGAGAGCAAGGCAA 188

      265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleG 282
          |||::::::|||   |||   :::::::::   |||   ||||||:::
```

```
      189 TGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGAATTCAAGCTGC 238

      282 IuAspSerALaGluValAlaArqLeuTrpGlyIleArqLysAsnArqPro 298
            |||:::   :::||||||||||||||||||||:::||||||||||:::
      239 TGGATGCAGAGGAGGTGGCCCGGCTGTGGGGACTACGCAACGAACAAGACC 288

      299 AlaMetAsnTyrAspLysLeuSerArqSerIleArqGluTyrTyrLysLy 345
            ||||||||||||||||||||||||||||:::::|||   ||||||   ||
      289 AACATGAATTACGACAAGCTCAGCCGGCCCTTGCGGTACTACTATGACAA 338

      315 sGlyIleIleArqLysProAspIleSerGluArqLeuValTyrGluPheV 332
          |:::|||||||||||||   :::   :::||:::::::|||||:::||||
      339 GAAcATCATCCGCAAGGTGAGC...GGCCAGAACTTCGTCTACAAGTTTG 385

      332 aI 332
          ||
      386 TG 387

seq_name: qb_est89:BF584991

Seq_documentation_block:
LOCUS       BF584991       861 bp    mRNA             EST       12 DEC 2000
DEFINITION  602070888F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4218741 5',
            mRNA sequence.
ACCESSION   BF584991
VERSION     BF584991.1  GI:11658709
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 861)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9799  row: j  column: 22
            High quality sequence stop: 530.
FEATURES             Location/Qualifiers
     source          1..861
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4218741"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: colon; Vector: pCMV.SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      186 a    318 c    248 g    109 t
ORIGIN


alignment scores:
            Quality:  223.50              Length:     144
              Ratio:  2.304                 Gaps:       8
 Percent Similarity:  67.361   Percent Identity: 45.833

alignment block:
 US-09-126-945B-2 x BF584991    ..

 Align_seq_1/1  to BF584991  from 1  to 861

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
```

```
          ||||||||||||||||||||||||||||:::|||||||||||:::   |||||
      421 ATGGGCAGTGCCAGCCCAGGCCTGAGCAACGTGTCCCCGGTGCCTGCC 470

       17 uLeuProProAspThrValSerArqThrGlyLeuGluLysAlaAlaAlaG 64
          ||||   |||||:::::   |||||||||   |||:::||||||||:::|
      471 ACTGTACCCAGATGTGGCACCAGAACAGGCCGGAGCAGCAGCATCAG 520

       64 lyAlaValGly..LeuGluArqArqAspTrpSerProSerProPro.Ala 49
          |||||:::|||  :::::::||||||   |||||||||||||||||| |||
      521 GAGCAATGGGCCCATGAGACGCAGGACATGGAGTCCTAGTCCACCACGCC 570

       50 ThrProGluGluGlyLeuSerAlaPheTyrLeuSerTyrPheAspMet.. 65
          :::|||||||||||||      ::::::      ::::::   |||:::
      571 AGCCCTGAGCAGGGC...CATGTCTGGCTAACTACACACACTGACTTCAA 617

       66 .LeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSer 81
          :::   |||:::|||||||||   :::|||||||:::||| |||   :::
      618 CATGGATCCCGACGATAGCAGCAGGGTCGCCAAAGTCCCGAGGCCGTGCC 667

       82 SerArqGluGluProProGluGluProGluGlnCysProValIleAspSe 98
          :::   ::::::   |||||||||||||||||   |||   :::   ::
      668 GGC...GACGACCAACCGGAAGAACCCGAGCAGGGTCCAGAAATGAAGA 714

       98 nGlnAla.........ProAlaGlySerLeuAspLeuValProGlyGlyL 112
          :::|||          |||
      715 CAAGGCTGGGACCAGCCC.............................. 744

      112 euThrLeuGlnGluHisSerLeuGlnGlnValGluSerMetValValGly 128
                :::|||:::|||||          |||:::       :::|||
      744 .......GATGAACAATCACTAGCAGCGGAGCAACAGCCGGGGGGTGGC 775

      129 GluValLeuLysAspIleGlyThrAla 147
              |||||      :::|||||||
      776 ACAGGCCTGAAAGCCACCGACACGGCA 802

seq_name: qb_est96:BG069473

seq_documentation_block:
LOCUS       BG069473       786 bp    mRNA             EST       26 JAN 2001
DEFINITION  H3075H09 3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3075H09 3', mRNA sequence.
ACCESSION   BG069473
VERSION     BG069473.1  GI:12551965
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   Bargot,T., Peckham,G.E., Qian,Y., Lim,M.K., Jaradat,S.A., Emaka
            ,I.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other ESTs: H3075H09.5
            Contact: George J. Karaul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdnaclusun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://www.grc.nia.nih.gov/branches/lmg/lpg/nia_clone.html for details.
            Plate: H3075  row: H  column: 09
            Seq primer:  21M13 Forward
            High quality sequence stop: 786
            POLYA Yes.
FEATURES             Location/Qualifiers
     source          1..786
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="H3075H09"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /sex="Clones arrayed from a variety of cDNA libraries"
```

```
                     /dev_stage "Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host "DH10B"
                     /note "Vector: pSPORT1; Site_1: SalI; Site_2: NotI. This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo cDNA libraries (including preimplantation stage
                     embryos from unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos, extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad) and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with oligo(dT
                     )-NotI primers. References include: (1) Genome-wide
                     expression profiling of mid-gestation placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray, 2000,
                     Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                     Large-scale cDNA analysis reveals phased gene expression
                     patterns during preimplantation mouse development, 2000,
                     Development, 127: 1737-1749; (3) Genome wide mapping of
                     unselected transcripts from extraembryonic tissue of
                     7.5-day mouse embryos reveals enrichment in the t-complex
                     and under-representation on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT       158 a    187 c    184 g    258 t
ORIGIN


alignment_scores:
            Quality:  223.00            Length:    205
              Ratio:  1.922               Gaps:      5
 Percent Similarity:  56.585   Percent Identity: 27.805

alignment_block:
 US-09-126-945B-2 x BG069473/rev  ..

 Align seq 1/1  to reverse of: BG069473  from: 1  to: 786

     128 GlyGlnValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleTh 177
         ||:::      :::|||:::      ::::::     :::   :::
     726 GGCCAATGCAGCAGTGACCTTTTCCAGTCCGCACACAATGTCATTGTCAA 677

     144 rAlaAspProMetAspTrpSerProSerAsnValGluLysIrpLeuLeu 193
          ::::::      |||   |||      |||:::              1
     676 GACTGAACAAACCGATCCTTCCATCATGAACACA.............I 642

     161 rpThrGluHisGlnTyrArqLeu...ProProMetGlyLysAlaPheGln 176
         ||   |||:::::::|||        |||      |||::::::   :::
     641 GGAAAGAAGAAAACTATCTCTATGATCCCAGCTATGGTAGCACAGTAGAT 592

     177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGlnGlnGlnPheArqGl 193
         |||   ::::||   ::::||      :::   :::::
     591 CTGTTGGACAGTAAGACTTCTGCCGGGCTCAGATCTCCATG......... 550

     193 nArqSerProLeuGlyGlyAspValLeuHisAlaHisGluAspIleTrpS 213
                              :::      :::||||||   :::   :
     549 .....................ACAACCTCCAGTCACCTTCCAGTTGCAG 522

     210 ysSerAlaAlaTrpMetLysGlnArqThrSerProGlyAlaIleHisTyr 226
         ::|||           ||||||:::      |||         |||
     521 AGTCACCTGATATGAAAAAGGAGCAAGACCACCCTGTAAAGTCCCAC... 475

     227 CysAlaSerThrSerGluGluSerTrpThrAspSerGlyValAspSerSe 246
                                                  :::::
     474 ..........................................ACCAAAAA 467

     243 rCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuL 260
         :   :::   :::      ||||||||:::|||:::::::::::::::|||
     466 GCACAACCCAAGAGGCACTCACTTATGGGAGTTCATCCGAGACATTCTCT 417

     260 euLysProHisSerTyrGlyArqPheIleArqIrpLeuAsnLysGluLys 276
         ||:::|||   :::          :::|||:::|||   ::::::::::::
     416 TGAGCCCAGACAAGAACCCAGGACTGATCAAATGGGAAGACCGTTCGGAA 367

     277 GlyIlePheLysIleGluAspSerAlaGlnValAlaArqLeuTrpGlyLl 293
         |||||||||:::::::      |||      |||||:::|||||||||
     366 GGCATCTTCAGGTTCCTGAAGTCAGAAGCTGTGGCTCAGCTGTGGGGGAA 317

     293 rArqLysAsnArqIrpAlaMetAsnTyrAspLysLeuSerArgSerIleA 310
          :::|||||||:::   :::|||:::|||:::|||||||||||||:::::|
     316 AAAGAAAAATAACAGTAGCATGACATACGAGAAGCTCAGCCGGGCTATGA 267

     311 rgGlnTyrTyrLysLysGlyIleIleArqLysProAspIleSerGlnArq 326
         ||   |||||||||:::   |||:::::::::   |||   :::::|||
     266 CATATTACTACAAACAGAAATCCTGGAACGTGTGGAT...GGACGACGG 220

     327 LeuValTyrGlnPhe 331
         |||||||||:::|||
     219 CTGGTCTATAAGTTT 205

seq_name: qb_est88:BF476482

seq_documentation_block:
LOCUS        BF476482       627 bp    mRNA             EST       05-DEC-2000
DEFINITION   naa2%07.xl NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:4256020 3'
             similar to TR:O95238 O95238 ETS TRANSCRIPTION FACTOR PDEF. ;, mRNA
             sequence.
ACCESSION    BF476482
VERSION      BF476482.1  GI:11547309
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 627)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael K.
             Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: M. Bento Soares, Ph.D.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL, send email to:
             info@image.llnl.gov
             Seq primer: -40UP from Gibco
             High quality sequence stop: 464.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:4256020"
                     /clone_lib "NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host "DH10B"
                     /note "Organ: prostate; Vector: pT7T3D Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clones IDs
                     986608-986759, 1101192-1101959, and 1217928-1220415).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       106 a    176 c    213 g    132 t
ORIGIN


alignment_scores:
            Quality:  222.00            Length:     51
              Ratio:  4.723               Gaps:      0
 Percent Similarity:  92.157   Percent Identity: 86.275
```

```
alignment block:
 US-09-126-945B-2 x BF476482/rev

 Align seq 1/1  to reverse of: BF476482  from: 1  to: 627

      285 AlaGlnValIleArgLeuTrpGlyIleArgLysAsnArgProAlaMetAs 401
          ||||||||||||||||||||||   |||      :::   ||||||||||| 
      627 GCACAGGTGGCCCGGCTGGACCATCAGCTAAACCTGTCCCGCAATGAA 578

      401 nTyrAspLysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleI 418
          ||||||||||:::||||||||||||||||:::|||||||||||||||||| 
      577 CTACGACAAAGTGAGCGCGCTCCATCCGCATGATTACAAGAAGGGCATCA 528

      418 IleArgLysProAspIleSerGlnArgLeuValTyrGlnPheValHisPro 434
          |||||||||||||||||||||||||||||||||||||||||||||||||| 
      527 TCCGGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACCCC 478

      435 Ile 435
          ||| 
      477 ATC 475

Seq name: qb_est54:AW085653

Seq documentation block:
LOCUS       AW085653       479 bp    mRNA            EST       02 JUN 2000
DEFINITION  uf84b02.y1 Soares mammary gland NMLMG Mus musculus cDNA clone
            IMAGE:1528779 5', similar to TR:Q92ZK9 Q92ZK9 E74-LIKE FACTOR 5., ,
            mRNA sequence.
ACCESSION   AW085653
VERSION     AW085653.1  GI:8180094
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:944879
            Seq primer:  40RP from Gibco
            High quality sequence stop: 472.
FEATURES             Location/Qualifiers
     source          1..479
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1528779"
                     /clone_lib="Soares mammary gland NMLMG"
                     /sex="female (Jackson)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       140 a    100 c    149 g    100 t
ORIGIN


alignment scores:
            Quality:  221.00            Length:    119
              Ratio:  2.563               Gaps:      3
 Percent Similarity:  69.748   Percent Identity: 49.496
```

```
alignment block:
 US-09-126-945B-2 x AW085653  ..

 Align seq 1/1  to: AW085653  from: 1  to: 479

      217 GluArgThrSerProGlyAlaIleHisTyr.............CysAlaSe 229
          |||:::::::::   |||        |||              |||   :: 
       88 GAACGAGACCAAGACTGGCATCAAAGACTATGCTGATTCCAGCTGCTTGAA 137

      229 rThrSerGluGluSerProThrAspSerGluValAspSerSerCysSerG 246
          :|||||       :::      ::::::::      |||      |||  
      138 AACAAGT......GGCATCAAGAGTCAAGACTGTCACAGCCGAACAAGCC 181

      246 lyGlnProIleHisLeuTrpGluPheLeuLysGluLeuLeuLeuLysPro 262
            |||       |||||||||||:|||||||||||||||||||||:|||  
      182 TCCAAAGTTCTCACCTGTGGGAATTTGTCAGAGACTTGCTGCTGTCCCCT 231

      263 HisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePh 279
          ::::::      ::::::::::|||   ::::::|||:::|||||||  
      232 GAAGAGAACTGTGGCATCCTGGAATGGCAAGACAGGGAGCAGGGCATTTT 281

      279 eLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysA 296
          |:::::      |||       ::|||||   ||||||:  
      282 CCGAGTGGTTAAGTCAGAAGCCTGGCAAAGATGTGGGGACAAAGGAAGA 331

      296 snArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyr 312
          :::::      |||:::|||:::|||||||||||||:::::|||   ||| 
      332 AGAATGACAGGATGACGTACGAGAAGCTGAGCCGAGCCCTGAGATACTAC 381

      313 TyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTy 329
          |||||   ||||||||||||||   |||       :::|||||||||||  
      382 TATAAAACGGGAATCATGGAGCGGCTGACCCCGGAGGTTAGTGTA 425

      329 rGlnPhe 331
          :::|||  
      426 CAAATTT 432

Seq name: qb_est87:BF443190

Seq documentation block:
LOCUS       BF443190       482 bp    mRNA            EST       01 DEC 2000
DEFINITION  260593 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF443190
VERSION     BF443190.1  GI:11503282
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Keaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4466
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and all trimmed with phred
            v0.980904.e. Vector identified by cross_match with the  minscore 18
            and  minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 92  row: K  column: 21
            Seq primer: ATTIAGGIGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..482
                     /organism="Sus scrofa"
```

```
                     /db_xref "taxon:9823"
                     /clone_lib "MARC 2PIC"
                     /tissue_type= "pooled"
                     /lab_host "DH10B"
                     /note "Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT       101 a    135 c    142 g    104 t
ORIGIN


alignment_scores:
            Quality:  221.00             Length:    120
              Ratio:  2.600                Gaps:      6
 Percent Similarity:  70.833   Percent Identity: 48.333

alignment_block:
 US-09-126-945b-2 x BF443190   ..

 Align seq 1/1  to: BF443190  from: 1  to: 182

     215 MetTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...Ph 279
         |||   :::   :::|||||||      :::          ||||||||||
     134 ATGATGTCTGCAATCGTCCCCGAGAAATTTGAAG......GTGACATCAAG 177

     231 rGluGluSerTrpThrAspSerGluValAspSerSerCys6et...Gly3 247
         |:::::::::|||      ||||||   |||          |||   |||
     178 CAGGAAGGGGTGGCAGCCTTCCGAGAGGGACCACCCTACCAGCGTCGGGG 227

     247 InProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHis 266
         |||   :::|||||||||||||||||      ||||||     |||
     228 TGCTTTC/CAGCTGTGCCAGTTTCTGCTGGCCCTGCTGGATCACCCC... 273

     264 SerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...Ph 279
         :::   ::::::|||||   |||   :::       :::|||:::   ||
     274 ACAAATGCCCACTTCATTGCCTGGACCGGC......CGGGGATGGAGTT 317

     279 eLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysA 296
         |||:::   :::       :::|||||||||||||||||||||:::|||
     318 CAAACTAATTGAGCCTGAGCAGGTCGCCAGGCTCTGGGGCATCCAGAAGA 367

     296 snArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyr 316
         ||||||||||||||||||||||||||||||||||||:::|||   |||
     368 ACCGGCCAGCTATGAACTATGACAAACTGAGCCGCTCACTCCGGTACTAC 417

     317 TyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTy 329
         |||::: ||||||||||:::::||      :::::|||   |||||
     418 TATGAGAAAGGCATCATGCAGAAGGTG...GCCGGCGAGCGTTACGTGTA 464

     329 rGlnPheVal 332
         |:::|||||
     465 CAAGTTGTG 474

seq_name: qb_est11:W98560

seq_documentation_block:
LOCUS       W98560          555 bp    mRNA            EST       16-JUL-1996
DEFINITION  ma14e04.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:423774 5' similar to qb:M25269 PROTEIN E1K 1 (HUMAN);
            qb:X87257 M.musculus mRNA for elk1 protein (M.MUS).; mRNA sequence.
ACCESSION   W98560
VERSION     W98560.1  GI:1428471
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:258326
            Seq primer: ETPrimer
            High quality sequence stop: 340.
FEATURES             Location/Qualifiers
     source          1..555
                     /organism "Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:423774"
                     /clone_lib "Soares mouse embryo NbME13.5 14.5"
                     /sex "unknown"
                     /tissue_type "embryo"
                     /dev_stage "13.5 14.5dpc total fetus"
                     /lab_host "DH10B"
                     /note "Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTT
                     T 3']; on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT       106 a    157 c    183 g    109 t
ORIGIN


alignment_scores:
            Quality:  221.00             Length:    156
              Ratio:  2.255                Gaps:      6
 Percent Similarity:  62.821   Percent Identity: 47.821

alignment_block:
 US-09-126-945B-2 x W98560   ..

 Align seq 1/1  to: W98560  from: 1  to: 555

     179 AlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSe 195
         |||:::::   |||      :::::::::   ||||:|||
      24 GCGGCTTCTGGTTGCTGCTTCTGCTGCTGCGCCGCGGCCACGGCGCGCGCCT 73

     195 rProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerA 212
            |||:::                              |||   ::::
      74 TGGAACCGGGGCCC.....................IGGGGGGCGG 96

     212 laAlaTrpMet|,sSLArgHisCys:,GlyAlaIleHisTyrCysAla 228
         :::::        |||         |||||||||            |||
      97 TGGGGCCTCGTATGGAGCCCCCGCCCCCGGAGCTGCCACTGCCACCGCC 146

     229 ..SerThrSerGluGluSerTrpThrAspSerGluValAspSerSerCys 244
           :::|||   :::   ||||||      :::   |||:::::|||
     147 GCGCGACTCGCAGGTACTCCTGG......AATGGCGTGAGTGCTTCCCTA 190

     245 SerGlyGlnPro...IleHisLeuTrpGlnPheLeuLysGluLeuLeuLe 260
           ::|||   :::   |||||||||||||||   :::   |||||
     191 GTGATGGACCCATCGGTGACGCTGTGGCAGTTTCTGCTGCAG...CTTCT 237

     260 uLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysG 277
         |:::   ::::::   |||:::::|||   |||   :::::::::   |
```

```
      238 GAGAGAACAAGGTAATGGCTACATCATCTCCTGGACCTCACGGGATGGTG 287

      277 lysIlePheLysIleGluAspSerAlaGluValAlaArgLeuTrpGlyIle 293
          ||   |||||||:::   |||:::   :::|||||||||||||||||||:::
      288 GTGAGTTCAAGTTGGTGGATGCAGAGGAGGTGGCCCGGCTATGGGGACTG 337

      294 ArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleAr 340
          ||||||||||:::      |||||||||||||||||||||||||:::|||
      338 CGCAAGAACAAGACCAACATGAATTACGACAAGCTTAGCCGGAGCATTGAG 387

      310 qGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgL 327
          |   ||||||   |||:::||||||||||:::   :::   :::|||:::
      388 GTACTACTATGATAAGAATATCATCCGCAGAGTGAGC...GGCCAGAAGT 434

      327 enValTyrGlnPheVal 342
            ||||||:::   |||
      435 GTGTCTACAAGTGTGTG 451

seq_name: qb_est96:BG076722

seq_documentation_block:
LOCUS         BG076722       555 bp    mRNA            EST       26 JAN 2001
DEFINITION    H3005A08 5' NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
              H3005A08 5', mRNA sequence.
ACCESSION     BG076722
VERSION       BG076722.1  GI:12559291
KEYWORDS      EST.
SOURCE        house mouse.
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE     1  (bases 1 to 555)
  AUTHORS     Karaul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
              ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE       Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL     Unpublished (2001)
COMMENT       Other ESTs: H3005A08 3
              Contact: George J. Karaul
              Laboratory of Genetics
              National Institute on Aging/National Institutes of Health
              333 Cassell Drive, Suite 4000, Baltimore, MD 21224 6820, USA
              Email: cdna@lgsun.grc.nia.nih.gov
              This clone set has been freely distributed to the community. Please
              visit http://lgsun.grc.nia.nih.gov/cDNA/15K.html for details.
              Plate: H3005  row: A  column: 08
              Seq primer:  21M13 Reverse
              High quality sequence stop: 555
              POLYA No.
FEATURES             Location/qualifiers
     source          1..555
                     /organism "Mus musculus"
                     /strain "C57BL/6J"
                     /db_xref "taxon:10090"
                     /clone "H3005A08"
                     /clone_lib "NIA Mouse 15K cDNA Clone Set"
                     /sex "Clones arrayed from a variety of cDNA libraries"
                     /dev_stage "Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host "DH10B"
                     /note "Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo cDNA libraries (including preimplantation stage
                     embryos from unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos, extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad) and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with oligo(dT
                     ) Not primers.  References include: (1) Genome wide
                     expression profiling of mid gestation placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray, 2000,
                     Proc. Natl. Acad. Sci. U.S.A. 97: 9127-9132. (2)
                     Large scale cDNA analysis reveals phased gene expression
                     patterns during preimplantation mouse development, 2000,
                     Development. 127: 1737-1749; (3) Genome-wide mapping of
                     uncloned transcripts from extraembryonic tissue of
                     7.5 day mouse embryos reveals enrichment in the t-complex
                     and under representation on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT       178 a    121 c    134 g    122 t
ORIGIN


alignment scores:
              Quality:  220.50              Length:      87
                Ratio:  4.341                 Gaps:       2
 Percent Similarity:  75.862   Percent Identity: 49.425

alignment block:
 US-09-126-945B-2 x BG076722   ..

 Align seq 1/1  to: BG076722  from: 1  to: 555

      246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 262
          ||:::   |||:::|||||||:::||||||      ||||||
       24 GGAAACACAATTTACCTTGGGAGTTTTGCTGGCCCTGCTTCAGGACAA 73

      262 oHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 279
            :::     :::::|||:::|||   :::::|||||||||||||||
       74 AGCTACCTGTCCTAACGTACATCAACGTGGACGCAGCGAGAAAGGGCATTT 123

      279 heLysIleGluAspSerAlaGluValAla.ArgLeuTrpGlyIleArgLy 295
          |||:::   ||||||      |||::: ||||||||||||   :::||
      124 TTAAGCTGGTTGATTCTAAAGCCGTGTCCTAGATTGTGGGGGAAGACAA 173

      295 sAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnT 312
          |||:::|||   ||||||||||:::   :::::||:::::::|||   |
      174 AAACAAACCTGACATGAACTACGACAAGCTAAGCCGTAGCATCAGGTACT 223

      312 yrTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVal 328
          |||||:::::|||||:::   |||   :::   :::|||||||||||||
      224 ATTACCAAAGAGGTATTCTTGCAAAAGTGGAA...GGTCAGCGCTTGGTG 270

      329 TyrGlnPhe 337
          |||||||||
      271 TATCAGTTT 279

seq_name: qb_est100:BG394169

seq_documentation_block:
LOCUS         BG394169      1065 bp    mRNA            EST       12 MAR 2001
DEFINITION    602456455F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4578997 5',
              mRNA sequence.
ACCESSION     BG394169
VERSION       BG394169.1  GI:13287617
KEYWORDS      EST.
SOURCE        human.
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 1065)
  AUTHORS     [illegible]
  TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL     Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Tech./Robbin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
              Plate: LLCM1295  row: m  column: 14
              High quality sequence stop: 556.
FEATURES             Location/qualifiers
     source          1..1065
                     /organism "Homo sapiens"
```

```
                     /db_xref "taxon:9606"
                     /clone "IMAGE:4578997"
                     /clone_lib "NIH_MGC_16"
                     /tissue_type "retinoblastoma"
                     /lab_host "DH10B (phage-resistant)"
                     /note="organ: eye; Vector: pCMV7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Jing Bona
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (life technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT       281 a    426 c    305 g    153 t
ORIGIN


alignment_scores
            Quality:  220.50              Length:     121
              Ratio:  2.722                 Gaps:       5
 Percent Similarity:  66.942   Percent Identity: 40.496

alignment_block:
 US-09-126-945B-2 x BG394169   ..

 Align seq 1/1  to: BG394169  from: 1  to: 1065

      226 SerProGlyAlaIleHisTyrProAla              SerThrArgI 247
          :::|||   :::   :::::|||                  :::|||   ::
      106 GCCCCCGGAGCTGCCAACATTGCCAACGCCACCGCCACGCTACACACAG 155

      232 uGluSerTrp............ThrAspSerGluValAspSerSerCysS 245
          :      |||            |||         :::|:   |||
      156 GTACCCCTGGGATGGCGTGAGCACTCCCCCAGCGATGGACCCATCT.... 201

      245 erGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLys 291
                :::   |||||||||||||||   :::   ||||||:::
      202 ..........CTGACCCTCTGGCAGTTTCTGCTGCAG...GTGCTGAGA 242

      292 ProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGly|| 338
          ::::::   ||||:::::|||   |||   :::::::::   |||
      243 GAGCAAGGCAATGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGA 287

      339 ePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgL 385
           ||||||:::   |||:::   :::|||||||||||||||||||:::|||
      288 ATTCAAGCTGGTGGATGCACAGGAGGTGGCCCGGCTGTGGGGGCTACGCA 337

      386 ysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgIl 431
          ||||:::      ||||||:|||||||||||||||||:::::|||
      338 AGAACAACGCCAACATCAATCCACCACAACGTCAGCCCGCCTTGCGGTAC 387

      432 TyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVa 478
          ||||||   |||:::|||||||||||   :::   ::::|||||:::|
      388 TACTATGACAAGAACATCATCCGCAAGGTGAGC   GGCCAGAACTTCGT 434

      428 IlyrGlnPheVal 332
          ||||:::::::|||
      435 CTACAAGTACGTG 447

seq_name: qb_est78:BE764557

seq_documentation_block:
LOCUS        BE764557      424 bp     mRNA            EST       19-SEP-2000
DEFINITION   PM1-NT0073-230500-002-c09-NT0073 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BE764557
VERSION      BE764557.1  GI:10194481
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 424)
  AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.R.,
             Brumstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE    20202668
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/cgi-bin/db/mod_ptnt.pl?id=NT0073 230
             500-002-c09&t3_2000-05-23&t4_1)
             Seq primer: puc 18 forward
             High quality sequence start: 6
             High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1..424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0073"
                     /dev_stage="Adult"
                     /note="organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716    Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       104 a    105 c     92 g    124 t
ORIGIN


alignment_scores
            Quality:  219.50              Length:      90
              Ratio:  4.092                 Gaps:       4
 Percent Similarity:  78.889   Percent Identity: 53.333

alignment_block:
 US-09-126-945B-2 x BE764557/rev   ..

 Align seq 1/1  to reverse of: BE764557  from: 1  to: 424

      211 CysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuL 260
          |||   ||||||      ||| ||||||||||||||   :::::|||
      409 TGCTATGGACAGTGCTATCACCCTGTGGCAGTTCCTTCTTCAGTTCCTGC 360

      260 euLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLys 276
            ||||||:::::      :::::|||   |||   :::::::::
      359 AGAAGCCTCAGAAC...AAGCACATGATCTGTGGACCCTAAATGAT... 316

      277 GlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIl 293
          |||   ||||||:::   :::::   :::|||||||||||||||||
      315 GGGCAGATTTAAGTTTTGCAGCAAAAGAC:TGGCTCGTCTCTGGGGGATT 266

      293 cArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleA 310
          ||||||||||:::|||   |||||||||||||||||||||||:::::|
      265 TCGCAAGAACAAGCCTAACATGAATTATGACAAACTCAGCCGAGCCCTCA 216

      310 rgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGluArg 326
          ||   ||||||   |||:::|||||:::|||   :::   :::||||::
      215 GATACTATTATGTAAAGAAIATCAICAAAAAAGTGAAT...GCTCAGAAG 169

      327 LeuValIyrGlnPheVal 332
          :::||||||:::||||||
```

```
     168 TTTGTGTACAAGTTTGTC 151

Seq name: qb|est29:AL560651

Seq documentation block:
LOCUS       AL560651      968 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL560651 LTI NFL010 B02 Homo sapiens cDNA clone CS0DL003Y101 5
            prime, mRNA sequence.
ACCESSION   AL560651
VERSION     AL560651.1  GI:12907320
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 968)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope  Centre National de Sequencage
            BP 191 91006 EVRY cedex  France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..968
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "CS0DL003Y101"
                     /clone_lib "LTI_NFL010_B02"
                     /sex "male"
                     /tissue_type "B cells from Burkitt lymphoma"
                     /note "Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies, Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8471
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT       195 a    320 c    297 g    155 t      1 others
ORIGIN


Alignment scores:
            Quality:  218.00            Length:    144
              Ratio:  2.496               Gaps:      7
 Percent Similarity:  63.194   Percent Identity:  48.194

Alignment block:
 US-09-126-945B-2 x AL560651  ..

 Align seq 1/1  to: AL560651  from: 1  to: 968

     203 HisAlaHisLeu.....AspIleTrpLysSerAlaAl 214
         |||  ||||||          :::  |||  ::::::::
      19 CACCACCACCTCCACCGCCGCCTCGGAACCCAGGCCTGGGGGGCCGTGGG 68

     215 rTrpMetLysGluArgthrSerProGlyAlaIleHisTyrCysAlaSer1 230
         :    |||       ||||||||||   :::   :::   :::|
      69 GCCGCGTATGGAGCCCCCGCCCCCCGGAGCTGCCAACATGCCAACGCCA 118

     231 hrSer.GluGluSer.....TrpThrAspSer.....Gl 239
         ||::: :::   :::        |||   ||:::
     119 CCGCCACGCTACACACAGGTACCCCTGGGATGGCGTGACACTCCCCCAGC 168

     240 nValAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuL 256
         :::|||   |||          :::   ||||||||||   |||:
     169 GATGGACCCATCT.....GTGACGCTGTGGCAGTCTCTGC 204

     257 ysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgIpLeu 272
         :::::   |||||:::   ::::::   |||::::::|||   |||

     204 GGCAG...CTGTGACAGAGCAAGGCAATGGCCACATCATCTCCTGGACT 250

     273 AsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaAr 289
         :::::::::   |||   ||||||:::   |||:::   :::||||||||
     251 TCAACGGGATGGTGCTGAATTCAAGCTGGTGGATGCAGAGGAGGTGGCCCG 300

     289 qLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuS 305
         ||||||||||:::||||||||:::      ||||||   ||||||||||
     301 GCTGTGGGGGCTACGCAAGAACAAGACCAACATGAATMACGACAAGCTCA 350

     306 crArgSerIleArgGluTyrTyrLysLysGlyIleIleArgLysProAsp 322
         ||||::::::|||   ||||||   |||:::||||||||||   :::
     351 GCTGGGCCTTGCGGTACTACTATGACAAGAACATCATCCGCAAGCTGAGC 400

     323 IleSerGlnArgLeuValTyrGlnPheVal 342
         :::|||::::::||||||:::|||||
     401 ...GGCCAGAAGTTCGTCTACAACGTTTGTG 427

Seq name: qb|est45:AW280428

Seq documentation block:
LOCUS       AW280428      562 bp    mRNA            EST       04-JAN-2000
DEFINITION  tj40d01.y1 zebrafish adult brain Danio rerio cDNA 5' similar to
            TR:Q9YHW7 Q9YHW7 ETS DOMAIN PROTEIN ;, mRNA sequence.
ACCESSION   AW280428
VERSION     AW280428.1  GI:6668977
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 562)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,E., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,K., Ritter,E.,
            Kohn,S., Shin,I., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other ESTs: tj40d01.x1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: John Nhai. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrum@Primarbatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Zebrafish identity (p value greater than 1e-99) found to:
            gi|2439441|gb|AA566810|AA566810 ZF A196 zebrafish adult Danio rerio
            cDNA 5'
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 491.
FEATURES             Location/Qualifiers
     source          1..562
                     /organism "Danio rerio"
                     /db_xref "taxon:7955"
                     /clone_lib "zebrafish adult brain"
                     /sex "mixed male and female"
                     /tissue_type "brain"
                     /dev_stage "adult"
                     /lab_host "E. coli DH10B"
                     /note "Vector: pZIPLOX; Site_1: NotI; Site_2: SalI;
                     Original library was constructed in lambdaZIPLOX.  Mass
                     excision of the cDNA library was performed to yield
```

```
                    pZIPLoX plasmids. Insert check was done in original
                    library."
BASE COUNT       149 a    147 c    148 g    147 t      1 others
ORIGIN


alignment_scores:
            Quality:  217.50              Length:     89
              Ratio:  4.398                 Gaps:      4
 Percent Similarity:  75.294   Percent Identity: 56.471

alignment_block:
 US-09-126-945B-2 x AW280428   ..

 Align seq 1/1  to: AW280428  from: 1  to: 562

      249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 295
          ::::::!!!!!!!!!!!!!!!!!!      !!!!!!      !!!   !!!
      248 CTGCAACTCTGGCAGTICCTGGTCACTCTGTIGGATGACCCG...TCCAA 294

      295 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLysI 287
           !!!:::!!!!!!   !!!   :::       :::!!!:::   !!!!!!
      295 CGGGCACTTCATCACTTGGACTGGC.....CGTGGGCTTGAGTTTAAAC 338

      288 leGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
          ::   :::       :::!!!!!!!!!   !!!!!!!!!:::!!!!!!!
      339 TCATCGAACCAGAGGAGGTTGCTCGCCGATGGGGCTATTCAGAAGAACAGG 388

      298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrTy 344
          !!!:!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!:::!!!   !!!!!!::
      389 CCTGCCATGAATTACGACAAGCTGAGCCGATCCATCCGTTACTACTATGA 438

      314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 431
          :!!!!!!!!!!:::::!!!          :::::!!!   !!!!!!   !
      439 GAAGGGCATCATGCAAAAGGTT...GCTGGAGAGCCGTACGTGTACAANI 485

      331 hoVal 332
          !!!!!
      486 TIGTC 490

seq_name: qb_est53:AW911816

seq_documentation_block:
LOCUS        AW911816       575 bp    mRNA            EST       25-MAY-2000
DEFINITION   ut39c06.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
             IMAGE:1513738 5' similar to TR:O70273 O70273 EST B-MYL AGING FACTOR
             ;, mRNA sequence.
ACCESSION    AW911816
VERSION      AW911816.1  GI:8077398
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 575)
  AUTHORS    NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Other_ESTs: ut39c06.x1
             Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:94059u
             Seq primer: -40RP from Gibco
             High quality sequence stop: 478.
FEATURES             Location/Qualifiers
     source          1..575
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1513738"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized. Library
                     was constructed by Benito Soares and M. Fatima Bonaldo. "
BASE COUNT       187 a    135 c    138 g    114 t      1 others
ORIGIN


alignment_scores:
            Quality:  217.00              Length:    204
              Ratio:  1.887                 Gaps:      5
 Percent Similarity:  56.373   Percent Identity: 27.451

alignment_block:
 US-09-126-945B-2 x AW911816   ..

 Align seq 1/1  to: AW911816  from: 1  to: 575

      128 GlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleTh 144
          !!!:::       :::!!!:::       ::::::      :::   :::
       18 GGCCAAIGCAGCAGIGACCTTITCCAGICCGCACACAATGTCATTGICAA 67

      144 rAlaAspProMetAspIrpSerProSerAsnValGlnLysTrpLeuLeuT 161
          ::::::      !!!   !!!       !!!:::             !
       68 GACTGAACAAACCCATCCTTCCATCATGAACACA..............I 102

      161 rpThrGluHisGlnTyrArgLeu...ProProMetGlyLysAlaPheGln 176
          !!   !!!:::::!!!          !!!      !!!:::::   :::
      103 GGAAAGAAGAAAACTATCTCTATGATCCCAGCTATGGTAGCACAGTAGAT 152

      177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGl 193
             !!!   :::!!!   :::!!!       :::   ::::::
      153 CTGTTGGACAGTAAGACTTTCTGCCGGGCTCAGATCTCCATG........ 194

      193 nArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpL 210
                                  :::      :::!!!!!   :::   :
      195 .....ACAACCTCCAGTCACCTTCCAGTTGCAG 222

      210 ysSerAlaAlaIrpMetLysGluArgThrSerProGlyAlaIleHisTyr 226
          ::!!!          !!!!!!!:::      !!!          !!!
      223 AGTCACCTGATATGAAAAAGCAGCAAGACCACCCTGTAAAGTCCCAC... 269

      227 CysAlaSerThrSerGlnGluSerTrpThrAspSerGluValAspSerSe 243
                                                      :::::
      270 .....ACCAAAAA 277

      243 rCysSerGlyGlnProIleHisLeuTrpGluPheLeuLysGluLeuLeuL 260
          :   :::   :::       !!!!!!!!!:::!!!!!!!!!!!!!!!!!!
      278 GTACAACCCAACACTCACTCACTTACTCCAGACTTCACCTCACACAAGTCT 327

      260 euLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLys 276
          !!:::!!!   :::          :::!!!:::!!!   ::::::::::::
      328 TGAGCCCAGACAAGAACCCAGGACTGATCAAATGGGAAGACCGTTCGGAA 377

      277 GlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIl 293
          !!!!!!!!!!:::::    !!!      !!!!!!:::!!!!!!!!
      378 GGCATCTTCAGGTTCCTGAACTCAGAAGCTGTGGCTCAGCTGTGGGGGAA 427

      294 eArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleA 310
          :::!!!!!!!!   :::!!!:::!!!:::!!!!!!!!!!!:::::::!
      428 AAAGAAAAATAACAGTAGCATGACATACGAGAAGCTCAGCCGGGCTATGA 477

      311 rgGlnTyrTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArg 326
          !!   !!!!!!!!!!:   !!!:::::::::   !!!   :::::!!!
      478 GATATTACTACAAACAAAAGGTGGAACTGTGTGGATCCGACGACGG 524
```

```
      327 LeuValTyrGln 330
          ||||||||||:::
      525 CTGCTCTATAAG 536

Seq name: qb_est27:AI991109

Seq documentation block:
LOCUS       AI991109      832 bp    mRNA            EST       27 OCT 1999
DEFINITION  wu86c03.x1 Soares_BreckQuacte_colon_NHCD Homo Sapiens cDNA clone
            IMAGE:2522308 3' similar to TR:Q70272 Q70272 EIC B Methanol
            ;, mRNA sequence.
ACCESSION   AI991109
VERSION     AI991109.1  GI:5838012
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 832)
  AUTHORS   NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer:  400P from Gibco
            High quality sequence stop: 456.
FEATURES             Location/Qualifiers
     source          1..832
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2522308"
                     /clone_lib="Soares_BreckQuacte_colon_NHCD"
                     /tissue_type="colonic mucosa from 3 patients with Crohn's
                     disease"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: colon; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site 1: Not I; Site 2: Eco RI; 3'
                     strand cDNA was primed with a Not I  oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCGGTCTTTTTTTTTTTTTTTTT 3'],
                     double stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT718 vector.  Library
                     went through one round of normalization. Tissue samples
                     provided by Dr. Brian BreckQuacte (Washington University,
                     dbreck@im.wustl.edu); colonic mucosa represents a range of
                     disease involvement from moderate to severe Crohn's
                     disease; samples include both perforating (fistulas) and
                     non perforating samples. Library constructed by Benito
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT       175 a    175 c    196 g    286 t
ORIGIN


Alignment scores:
            Quality:  216.00              Length:    128
              Ratio:  2.602                 Gaps:      2
 Percent Similarity:  64.844   Percent Identity:  53.594

Alignment block:
 US-09-126-945B-2 x AI991109/rev  ..

 Align seq 1/1  to reverse of: AI991109  from: 1  to: 832

      204 AlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGlnArgThrG 220
          :::||||||   :::   :::|||            |||||||:::
      561 AGTCACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCC 512

      220 rProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrA 237
          |||:::      |||
      511 CCCTGCCAACTGCCAC.............................. 456

      237 spSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGln 253
                ::::::   :::   :::      ||||||||||:::
      495 ..........ACCAAAAAGCACAACCCGAGACGGACTCACTTATGGCAA 457

      254 PheLeuLysGluLeuLeuLysProHisSerTyrGlyArgTheIleAr 270
          |||::::::::::::::|||||:::|||   :::         :::|||:::
      456 TTCATCCGCGACATCCTCTTGAACCCAGACAACGAACCCAGGATTAATAAA 407

      220 rTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnV 287
          :|||   :::::::::::::|||:::|||::::       |||     |
      406 ATGGCAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGCAG 357

      287 alAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAsp 303
          ||||||:|||||||||   :::|||||||:::   :::|||:::|||:::
      356 TGGCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTATGAA 307

      304 LysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLy 320
          ||||||||||||:::::|||   ||||||||||:::   |||:::::::::
      306 AACTTAGTCGCAGTTATCGCACATATACTACAAAGAGGAAAACTTCACAA 257

      320 sProAspIleSerGlnArgLeuValTyrGlnPhe 331
          :   |||   :::::|||||||||||||:::|||
      256 TGTGGATC...GGACGAAGACTGGTATATAAATTT 226

Seq_name: qb_est74:BE814959

Seq documentation block:
LOCUS       BE814959      616 bp    mRNA            EST       26 OCT 2000
DEFINITION  601140404F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3049752 5',
            mRNA sequence.
ACCESSION   BE814959
VERSION     BE814959.1  GI:9145115
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Life Tech/Research Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM94  row: g  column: 01
            High quality sequence stop: 615.
FEATURES             Location/Qualifiers
     source          1..616
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3049752"
                     /clone_lib="NIH_MGC_9"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: ovary; Vector: pCMV; Site 1: XhoI; Site 2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size selected >500bp for average
                     insert size 1.8kb.  Library constructed by Life Tech in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  CAP cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT       121 a    178 c    182 g    135 t
ORIGIN


Alignment scores:
```

```
            Quality:  215.50              Length:     85
              Ratio:  3.265                 Gaps:      4
 Percent Similarity:  77.647   Percent Identity: 55.294

alignment block:
 US-09-126-945B-2 x BE414959  ..

 Align seq 1/1  to: BE414959  from: 1  to: 616

      249 IleHisLeuIrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
          ::::::||||||||||||||||      ||||||     |||   :::
      119 CTGCAGCTGTGGCAATTTCTGGTGGCCTTGCTGGATGACCCA...ACAAA 165

      265 rGlyArqPheIleArqIrpLeuAsnLysGluLysGlyIle...PheLysI 281
          :::::::|||||   |||   :::      :::|||:::   ||||||:
      166 TGCCCATTTCATTGCCTGGACGGGC.....CGGGGAATGGAGTTCAAGC 209

      281 leGlnAspSerAlaGlnValAlaArqLeuIrpGlyIleArqLysAsnArq 297
          ::   :::      :::|||||||||||||||||||||:::|||||:|||
      210 TCATTGAGCCTGAGGAGGTCGCCAGGCTCTGGGGCATCCAGAAGAACCGG 259

      298 ProAlaMetAsnIyrAspLysLeuSerArqSerIleArqGlnTyrIyrLy 344
          |||||||||||||||||||||||||||||||||:::|||   |||||::
      260 CCAGCCATGAATTACGACAAGCTGAGCCGCTCGCTCCGATACTATTATGA 309

      344 sLysGlyIleIleArqLysProAspIleSerGlnArqLeuValIyrGlnP 331
          :|||||||||:::::::|||         :::::|||   |||||||:::|
      310 GAAAGGCATCATGCAGAAGCTG...GCTGGTGAGCGTTACGTGTACAACCT 356

      331 heVal 332
          |||||
      357 TTGTG 361

seq_name: gb_est46:AW411287

seq_documentation_block:
LOCUS       AW411287      640 bp    mRNA            EST       29-JUN-2000
DEFINITION  tb11d10.y1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2964546 3',
            mRNA sequence.
ACCESSION   AW411287
VERSION     AW411287.1  GI:6966828
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Plate: LLCM64  row: H  column: 19
            Seq primer: M13RPl reverse primer (ABI).
FEATURES             Location/qualifiers
     source          1..640
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2964546"
                     /clone_lib="NIH_MGC_17"
                     /tissue_type="rhabdoid, soft tissue"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      132 a    185 c    182 g    141 t
ORIGIN


alignment_scores:
            Quality:  215.50              Length:     85
              Ratio:  3.265                 Gaps:      4
 Percent Similarity:  77.647   Percent Identity: 55.294

alignment_block:
 US-09-126-945B-2 x AW411287  ..

 Align seq 1/1  to reverse of: AW411287  from: 1  to: 640

      249 IleHisLeuIrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
          ::::::|||||||:||||||      ||||||     |||   :::
      480 CTGCAGCTGTGGCAATTTCTGGTGGCCTTGCTGGATGACCCA...ACAAA 434

      265 rGlyArqPheIleArqIrpLeuAsnLysGluLysGlyIle...PheLysI 281
          :::::::|||||   |||   :::      :::|||:::   ||||||:
      433 TGCCCATTTCATTGCCTGGACGGGC.....CGGGGAATGGAGTTCAAGC 390

      281 leGlnAspSerAlaGlnValAlaArqLeuIrpGlyIleArqLysAsnArq 297
          ::   :::      :::|||||||||||||||||||||:::|||||:|||
      389 TCATTGAGCCTGAGGAGGTCGCCAGGCTCTGGGGCATCCAGAAGAACCGG 340

      298 ProAlaMetAsnIyrAspLysLeuSerArqSerIleArqGlnTyrIyrLy 344
          |||||||||||||||||||||||||||||||||:::|||   |||||::
      339 CCAGCCATGAATTACGACAAGCTGAGCCGCTCGCTCCGATACTATTATGA 290

      344 sLysGlyIleIleArqLysProAspIleSerGlnArqLeuValIyrGlnP 331
          :|||||||||:::::::|||         :::::|||   |||||||:::|
      289 GAAAGGCATCATGCAGAAGCTG...GCTGGTGAGCGTTACGTGTACAACCT 243

      331 heVal 332
          |||||
      242 TTGTG 238

seq_name: gb_est95:BF976822

seq_documentation_block:
LOCUS       BF976822      820 bp    mRNA            EST       22-JAN-2001
DEFINITION  602145626F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4305518 5',
            mRNA sequence.
ACCESSION   BF976822
VERSION     BF976822.1  GI:12444037
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1172  row: j  column: 15
            High quality sequence stop: 658.
FEATURES             Location/qualifiers
     source          1..820
```

```
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "IMAGE:4405518"
                     /clone_lib "NIH_MGC_62"
                     /tissue_type "melanotic melanoma, high MDR"
                     /lab_host "DH10B (T1 phage resistant)"
                     /note "organ: skin; Vector: pDNR-LIB (Clontech); Site 1:
                     SfiI (ggccgagctcggcc); Site 2: SfiI (ggccgattatggcc);
                     Double stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5' CACGGCCATTATGGCC 3' and 3' adaptor
                     sequence:  5' ATTCTACAGGCCGAGGCGGCCGACATG d(T30)BN 3'
                     (where B = A, C, or G and N = A, C, G, or T).  Average
                     insert size 1.7 Kb (range 0.7-3.8 Kb).  17.2% of
                     contained  inserts by PCR. This library was enriched for
                     full length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT       179 a    220 c    246 g    184 t
ORIGIN


Alignment scores:
            Quality:  215.50              Length:     85
              Ratio:  3.265                 Gaps:      4
 Percent Similarity:  77.647   Percent Identity: 55.294

Alignment block:
 US-09-126-945B-2 x BP976822   ..

 Align seq 1/1  to: BP976822  from: 1  to: 820

      249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLysProHisSerTy 295
          ::::::||||||||||||||||      ||||||      |||   :::
           77 CTGCAGCTGTGGCAATTTCTGGTGGCCTTGCTGGATGACCCA...ACAAA 123

      265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLys 311
           ::::::||||||   |||   :::      :::|||:::   ||||||:
      124 TGCCCATTCATTGCCCTGGACGGGC.....CGGGGAATGGAGTTCAAGC 167

      281 IleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 327
          ::   :::      :::||||||||||||||||||||||||:::|||||||||
      168 TCATTGAGCCTGAGGAGGTCGCCAGCCTCTGGGGCATCCAGAAGAACCGG 217

      298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyrTyrLy 344
          ||||||||||||||||||||||||||||||||||:::|||   ||||||:
      218 CCACCCATGAATTACGACAAGCTGAGCCGCTCGCTCCGATACTATTATGA 267

      314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 361
          :|||||||||:::::||          :::::|||   ||||||:::|
      268 GAAAGGCATCATGCAGAAGGTG...GCTGGTGAGCGTTACGTGTACAAGT 314

      331 heVal 332
          |||||
      315 TTGTG 319

Seq name: qb_est77:BE627411

Seq documentation block:
LOCUS       BE627411       655 bp    mRNA              EST       24-AUG-2000
DEFINITION  qt93109.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:3370025 3' similar to TR:O70273 O70273 ETS HOMOLOGOUS FACTOR
            ;, mRNA sequence.
ACCESSION   BE627411
VERSION     BE627411.1  GI:9907831
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 655)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other ESTs: qt93109.y1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty free through LENL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1079629
            High-quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1..655
                     /organism "Mus musculus"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:3370025"
                     /clone_lib "Soares_mammary_gland_NMLMG"
                     /sex "female (lactating)"
                     /tissue_type "mammary gland"
                     /lab_host "DH10B"
                     /note "Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       138 a    155 c    136 g    225 t      1 others
ORIGIN


Alignment scores:
            Quality:  215.00              Length:    128
              Ratio:  2.560                 Gaps:      2
 Percent Similarity:  65.625   Percent Identity: 54.375

Alignment_block:
 US-09-126-945B-2 x BE627411/rev   ..

 Align seq 1/1  to reverse of: BE627411  from: 1  to: 655

      204 AlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSe 220
          :::||||||   :::   :::|||          ||||||:::
      562 AGTCACCTTCCAGTGGCAGAGTCTCGTGATATGAAAAGGAGCAAGACCA 513

      220 rProGlyAlaIleHisTyrCysAlaSerThrSerGlnGluSerTrpThrA 267
           |||   :::   |||
      512 CCCTGIANGGTCCCAC................................ 497

      267 spSerGlnValAspSerSerCysSerGlyGlnProIleHisLeuTrpGln 253
                  ::::::   :::   :::      ||||||||||:::
      496 ..........ACCAAAAAGCACAACCCAAGAGGCACTCACTTATGGGAG 458

      254 PheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleAr 270
          |||:::::::::::::|||||:::|||   :::   :::   :::|||:::
      457 TTCATCCGAGACATTCTCTTGAGCCCAGACAAGAACTCAGGACTGATCAA 408

      270 qTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnV 287
          :|||   :::::::::::::|||||||||:::::      |||      |
      407 ATGGCAAGACCGTTCGGAAGGCATCTTCAGGTCCTGAAGTCAAAAGCTG 358

      287 aIleArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAsp 303
          ||||||:::|||||||||   :::|||||:::   :::|||||:::::
      357 TGGCTCAGCTGTGGGGGAAAAACAAAAATAACAGTAGCATGACATACAG 308

      304 LysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleIleArgLy 320
          ||||:||:|||::::::|||   ||||||||||:::   |||:::::::::
      307 AAGCTCAGCCGCTCCATCAGAATATTACTACAAACGAGAAATCATGGAACG 258

      320 sProAspIleSerGlnArgLeuValTyrGlnPhe 341
          :   |||   :::::|||||||||||||:::|||
      257 TGTGGAT...GGACGACGGCTGGTCTATAAGTTT 227

Seq name: qb_est38:AV405665
```

```
seq_documentation_block:
LOCUS        AV405665      823 bp    mRNA            EST       05-FEB-2000
DEFINITION   AV405665 Bombyx mori wing disk C108 5th instar day 3 larva Bombyx
             mori cDNA clone wdV40090 T3, mRNA sequence.
ACCESSION    AV405665
VERSION      AV405665.1  GI:6909753
KEYWORDS     EST.
SOURCE       domestic silkworm.
  ORGANISM   Bombyx mori
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
             ; Bombycoidea; Bombycidae; Bombyx.
REFERENCE    1  (bases 1 to 823)
  AUTHORS    Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
  TITLE      Bombyx mori cDNA
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Mita K
             Genome Research Group
             National Institute of Radiological Sciences
             Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
             Email: kmita@nirs.go.jp
             method:uni-directional, sequence direction:Sequenced from T3 primer
             (5' -> 3')
             Project '[illegible]
             Future Program in JSPS', see 'SilkBase',
             <http://www.[illegible]
FEATURES             Location/Qualifiers
     source          1..823
                     /organism="Bombyx mori"
                     /strain="C108"
                     /db_xref="taxon:7091"
                     /clone="wdV40090"
                     /clone_lib="Bombyx mori [illegible]
                     larva"
                     /sex="female/male mixed"
                     /tissue_type="wing disk"
                     /dev_stage="5th instar day 3 larva"
BASE COUNT      298 a    167 c    161 g    197 t
ORIGIN


alignment_scores:
            Quality:  214.00              Length:     83
              Ratio:    3.292               Gaps:      2
 Percent Similarity:   78.313   Percent Identity: 49.398

alignment_block:
 US-09-126-945B-2 x AV405665   ..

 Align seq 1/1  to: AV405665  from: 1  to: 823

     251 LeuTrpGlnPheLeuLysGluLeuLeuLeuLysPro...HisSerTyrGl 266
         ||||||::|||||||::::::|||||      |||   :::   |||
      61 TTGTGGGACTTCTTACAACAACTTCTGAATGACCCAACGCAAACATACAC 110

     266 yArgPheIleArgTrpLeuAsnLysGlnLysGlyIlePheLeuLeuA 313
         :::::::|||   |||   |||::::::   |||::|||||||||   |
     111 AAACIAIAITGCAIGCAAGAACAGAGACACIGGCGIATTIAAAATTGTTG 160

     283 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299
         ||   |||   :::|||::|||||||||||||||::|||||||:   :::
     161 ATCCTGCTGGACTAGCCAAATTGTGGGGAATTCAAAAAAATCACCTATCC 210

     300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
         |||||||||||||||||:::|||||:::::|||   ||||||::   ::
     211 ATGAACTACGACAAAATGTCCCGCGCCTTGAGATATTATCAGTGTTAA 260

     316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal 332
         :|||::::|||||   :::   :::::::|||      |||||||||::
     261 TATTTTGCGCAAAGICCAG...GGCGAAACACACTGTIACCAGTICTTA 306

seq_name: sb_est77:BE616829
```

```
seq_documentation_block:
LOCUS        BE616829      995 bp    mRNA            EST       24-AUG-2000
DEFINITION   601280609F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4621795 5',
             mRNA sequence.
ACCESSION    BE616829
VERSION      BE616829.1  GI:9898428
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 995)
  AUTHORS    NIH MGC Project
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM296  row: b  column: 04
             High quality sequence start: 43
             High quality sequence stop: 721.
FEATURES             Location/Qualifiers
     source          1..995
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4621795"
                     /clone_lib="NIH_MGC_49"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP cDNA
                     synthesis kit (Stratagene) and Superscript II RI (Life
                     Technologies)."
BASE COUNT      215 a    279 c    349 g    152 t
ORIGIN


alignment_scores:
            Quality:  214.00              Length:    124
              Ratio:    2.892               Gaps:      5
 Percent Similarity:   59.677   Percent Identity: 41.129

alignment_block:
 US-09-126-945B-2 x BE616829   ..

 Align seq 1/1  to: BE616829  from: 1  to: 995

     209 IrpLysSerAlaAlaIrpMetLysGluArg..................... 218
         ||||||::::::|||      :::   |||
       9 IGGAAGGGGGIGGCGAIAIGGGAGCCGCGICCCACGGAGCTGCAACAII 58

     219 ....IhrSerProGlyAlaIleHisTyrCysAlaSerThrSerGlnGlus 234
             |||   |||   |||   |||                   :::
      59 GCCAACGCCCACCGCACGCTAGACAC.................AGGIACC 90

     234 crIrp.............ThrAspSerGlyValAspSerSerCysSerGly 246
         |||             |||          :::|||   |||
      91 TCTGGGAIGGCGIGAGCACICCCCCAGCGGAIGGACCCAICI......... 131

     247 GluProIleHisLeuIrpGlnPheLeuLysGluLeuLeuLeuLysProHi 263
             :::   |||||||||||||||   :::   ||||||:::   ::
     132 .....GIGACGCIGIGGCAGIIICIGCGICAG...CIGCIGAGAGAGCA 172
```

```
      263 sSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePhe> 280
          ::::   |||::::::|||   |||   :::::::::   |||   ||||
      173 AGGCAATGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGAATTCA 222

      280 ysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsn 296
          ||:::   |||:::   :::|||||||||||||||||||||:::|||||||
      223 AGCTGCTGGATGCAGAGGCAGGTGGCCCGGCTGTGGGGGCTACGCAAGAAC 272

      297 ArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyrTy 313
          :::       |||||||||||||||||||||||||||:::::|||   |||||
      273 AAGACCAACATGAATTACGACAAGCTCAGCCGGGCCTTGCGGTACTACTA 322

      313 rLysLysGlyIleIleArgLys 320
          |   |||:::|||||||||||
      323 TGACAAGAACATCATCCGCAAG 344

Seq name: qb_est80:BE883422

Seq documentation block:
LOCUS       BE883422       893 bp    mRNA             EST       20-OCT-2000
DEFINITION  601509176F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3910699 5',
            mRNA sequence.
ACCESSION   BE883422
VERSION     BE883422.1  GI:10332198
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 893)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9726  row: q  column: 23
            High quality sequence stop: 667.
FEATURES             Location/Qualifiers
     source          1..893
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3910699"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: uterus; Vector: pCMV SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 2.1 kb.  "
BASE COUNT       180 a    286 c    276 g    151 t
ORIGIN


Alignment scores:
            Quality:  213.50              Length:     119
              Ratio:  2.773                 Gaps:       6
 Percent Similarity:  64.706   Percent Identity: 44.697

Alignment block:
 US-09-126-945B-2 x BE883422  ..

 Align seq 1/1  to: BE883422  from: 1  to: 893

      219 ThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerIl 235
          |||   |||   |||   |||                   :::      ||
       80 ACGCCACCGCACGCTACACAC..................AGGTACCCCTG 111

      236 p.............ThrAspSerGlnValAspSerSerCysSerGlyGlnP 248
          |             |||          :::|||   |||
      112 GGATGGCGTGAGCACTCCCCCAGCGATGCACCCATCT.............. 148

      248 roIleHisLeuTrpGln.PheLeuLysGluLeuLeuLeuLysProHisSe 264
            :::   ||||||||| ||||||   :::   |||||||::   :::::
      149 ..GTGACGCTGTGGCAGGTTTCTGCTGCAG...CTGCTGAGACAGCAAGG 193

      264 rTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysI 281
          :   |||::::::|||   |||   :::::::::   |||   |||||||
      194 CAATGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGAATTCAAGC 243

      281 leGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
          ::   ||:::   :::|||||||||||||||||||||:::|||||||||
      244 TGCTGGATGCAGAGGCAGGTGGCCCGGCTGTGGGGGCTACGCAAGAACAAG 293

      298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyrTyrLy 314
            |||||||||||||||||||||||||||:::::|||   |||||||
      294 ACCAACATGAATTACGACAAGCTCAGCCGGGCCTTGCGGTACTACTATGA 343

      314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
           |||:::|||||||||||||   :::   :::|||:::::|||||||:::|
      344 CAAGAACATCATCCGCAAGCTGAGC...GGCCAGAAGTTCGTCTACAAGT 390

      331 heVal 332
          |||||
      391 TTGTG 395

Seq_name: qb_est98:BG294409

Seq_documentation block:
LOCUS       BG294409       817 bp    mRNA             EST       21-FEB-2001
DEFINITION  602391518F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4503124 5',
            mRNA sequence.
ACCESSION   BG294409
VERSION     BG294409.1  GI:13055014
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 817)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10374  row: d  column: 05
            High quality sequence stop: 719.
FEATURES             Location/Qualifiers
     source          1..817
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4503124"
                     /clone_lib="NIH_MGC_94"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally; oligo dT primed.
                     Average insert size 3.8 kb. Library enriched for
                     full length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT       254 a    171 c    214 g    179 t
ORIGIN
```

```
alignment_scores:
            Quality:  213.00              Length:     199
              Ratio:  2.029                 Gaps:       7
 Percent Similarity:  52.764   Percent Identity: 29.648

alignment_block:
 US-09-126-945B-2 x BG294409   ..

 Align seq 1/1  to: BG294409  from: 1  to: 817

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIrpSe 151
          ||||||:::  ::::::  :::       ::::::|||||||||  ||
      199 GAGACTGTAGTGGAGGTGTCCACAGAAGAGTCTGAGCCGATGGATGCCTC 248

      151 rProSerAsnValGlnLysIrpLeuLeuIrpIhrGluHisGlnTyrArgL 168
          ||||                                          :
      249 GCCT.............................................A 253

      168 enProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCys 184
          :::||          :::   :::   :::      ||||::
      254 TCCCTACTTCACCAGACAGCCACGAACCAATGAAAAACAAAAAA...... 297

      185 AlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspVa 201
            ::::::::::::        ||||::::|||||:::::::::
      298 ...GTTGGCCGTAAACCAAAGACCAGCAGTCACCAGTTTCCAATGGG... 342

      202 lLeuHisAlaHisLeuAspIleIrpLysSerAlaAlaIrpMetLysGluA 218
              ::|||   |||   |||::|||         |||
      343 ....ICTCCTGAGTTAGGAATCATGAACAAAGCCAGAGAAGGAAAA.... 384

      218 rqIhrSerProGlyAlaIleHisTyrCysAlaSerIhrSerGluGluSer 234

      384 .................................................. 384

      235 IrpIhrAspSerGluValAspSerSerCysSerGlyGlnProIleHisLe 251
                                        |||:::     :::||
      385 ..............................GGAAACACAACCTATCT 401

      251 nIrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerIyrGlyArgP 268
          |||::::|||||  :::|||:|||      ::::::       ||||
      402 GTGGGAGTTTCTTTTAGATTGCTTCAAGATAAAAATACTTGTTCCAGGT 451

      268 heIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSer 284
          ::|||::|||   :::::|||||||||||||||||:::   ||||||
      452 ATATTAAATGGACTCAGACAGAAAAGGGCATATTCAAACTTGTGGACTCA 501

      285 AlaGlnValAlaArgLeuIrpGlyIleArgLysAsnArgProAlaMetAs 301
            ||::::::::|||||||   :::|||||::::|||   ||||
      502 AAGGCTGTCTGGAAGCTCTGGGGCAAGCATAAGAACAAACCAGATATGAA 551

      301 nIyrAspLysLeuSerArgSerIleArgGln.TyrTyrLysLysGlyIle 317
          |||:::  :::::|||:::::|||    ||||||:::::::|||.|| 
      552 CTATGAAACCATGGGGCGAGCGTTAAGATATTTACTACAAAGGGAATT 601

      318 IleArgLysProAspIleSer.GlnArgLeuValTyrGlnPhe 331
          :::  ||  :::   ::: ||||||||||||||||||||
      602 CTTGCAAAGGTTGAA...GGACCAGAGGCTTGTATATCAGTTC 641

seq_name: qb_est80:BE859344

seq_documentation_block:
LOCUS       BE859344       441 bp    mRNA            EST       29-SEP-2000
DEFINITION  UI-M-AQ0-aai-q-04-0-UI.r1 NIH_BMAP_MHI Mus musculus cDNA clone
            UI-M-AQ0-aai-q-04-0-UI 5', mRNA sequence.
ACCESSION   BE859344
VERSION     BE859344.1  GI:10375090
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromathi, Muridae, Murinae; Mus.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd, Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1..441
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-AQ0-aai-q-04-0-UI"
                     /clone_lib="NIH_BMAP_MHI"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pIT3B-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; the
                     NIH_BMAP_MHI library is a non normalized library
                     constructed from  mouse hippocampus. The tag is a string
                     of 5 nucleotides present  between the Not I site and the
                     oligo-dT track.  The library was  constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996. Tissue provided by Ms. Annie Novakovich,
                     Zivic Miller Laboratories."
BASE COUNT       112 a    112 c    116 g    101 t
ORIGIN


alignment_scores:
            Quality:  212.50              Length:      85
              Ratio:  3.269                 Gaps:        4
 Percent Similarity:  76.471   Percent Identity: 55.294

alignment_block:
 US-09-126-945B-2 x BE859344   ..

 Align seq 1/1  to: BE859344  from: 1  to: 441

      249 IleHisLenTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerIy 265
          :::::|||||||||||||||      ||||||    |||   |||
       29 CTCCAGCTCTGGCAGTTTTTGGTAGCTCTTCTGGATGACCCT...TCAAA 75

      265 rGlyArgPheIleArgIrpLeuAsnLysGluLysGlyIle...PheLys1 281
          ::::::::::::   |||   :::       ::::::::   ||||||
       76 TTCTCATTTCATTGCCTGGACTGGA......CGAGGCATAGGAATTTAAAC 119

      281 leGluAspSerAlaGlnValAlaArqLeuIrpGlyIleArqLysAsnArq 297
          ::   :::       :::|||||||||   ||||||||||:::|||||||||
      120 TGATTGAGCCCGAAGAGGTGGCCCGGCGTTGGGGCATTCAGAAGAACAGG 169

      298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 314
          ||||||||||||||||||||||||||||||||||:::|||   ||||:::
      170 CCGGCGATGAACTATGACAAACTTAGTCGTTCTCTCCGCCAATTATTATGA 219

      314 sLysGlyIleIleArgLysProAspIleSerGlnArqLeuValIyrGlnP 331
          :||||||||||:::::|||          :::::|||   ||||||:::|
      220 GAAGGGAATCATGCAAAAGGTG...GCTGGAGAAAGATACGTGTACAAAT 266

      331 heVal 332
          |||||
      267 TTGTG 271
```

```
Seq name: qb_est28:AL120951

Seq documentation block:
LOCUS       AL120951      446 bp    mRNA            EST       25-FEB-2000
DEFINITION  DKFZp762E043 ;) 76.2 (synonym: hmel2) Homo Sapiens cDNA clone
            DKFZp762E043 5', mRNA sequence.
ACCESSION   AL120951
VERSION     AL120951.1  GI:5926952
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     EST (Koehrer, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Koehrer K
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the Charite,
            Berlin/Germany) within the cDNA sequencing consortium of the German
            Genome Project.
            No s1 sequence available.
            This clone (DKFZp762E043) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..446
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp762E043"
                     /clone_lib="762 (synonym: hmel2)"
                     /tissue_type="melanoma (MeWo cell line)"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT       125 a    100 c    111 g    110 t
ORIGIN


alignment scores:
            Quality:  212.50            Length:      85
              Ratio:  3.269               Gaps:       4
 Percent Similarity:  76.471   Percent Identity: 55.294

alignment block:
 US-09-126-945B-2 x AL120951  ..

 Align seq 1/1  to: AL120951  from: 1  to: 446

     249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 295
         :::::|||||||||||||||||     ||||||     |||   |||
      68 CTTCAGCTCTGGCAGTTTTGGTAGCTCTTCTGGATGACCCT...TCAAA 114

     265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLysI 281
          :::::|||||   |||   :::      :::|||:::   ||||||:
     115 TTCTCATTTTATTGCCTGGACTGGT.....CGAGGCATGGAATTTAAAC 158

     281 IeGluAspSerAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArg 297
         ::   :::       :::||||||||||   ||||||||||:::||||||||
     159 TGATTGAGCCTGAAGAGGTGGCCCGACGTTGGGGCATTCAGAAAACAGG 208

     298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 344
         ||||||||||||||||||||||||||||||||||:::|||   |||||||:
     209 CCAGCTATGAACTATGATAAACTTAGCCGTTCATCTCCGCTATTACTATGA 258

     344 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
         :|||||||||:::::|||         :::::|||   ||||||:::|
```

```
     259 GAAAGGAATTATGCAAAAGCTG...GCTGGAGAGACATATGTCTACAAGT 305

     331 heVal 332
         |||||
     306 TTGTG 310

Seq name: qb_est76:BE553236

Seq documentation block:
LOCUS       BE553236      622 bp    mRNA            EST       15-AUG-2000
DEFINITION  nr43q06.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:3153082 5',
            similar to qb:LI0426 Mus musculus cts related protein 81 (MOUSE);,
            mRNA sequence.
ACCESSION   BE553236
VERSION     BE553236.1  GI:9817723
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1055848
            Seq primer: -40RP from Gibco
            High quality sequence stop: 446.
FEATURES             Location/Qualifiers
     source          1..622
                     /organism="Mus musculus"
                     /strain="C57BL/6N-A"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3153082"
                     /clone_lib="NCI_CGAP_Mam2"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="organ: mammary; Vector: pCMV SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT       162 a    153 c    167 g    140 t
ORIGIN


alignment scores:
            Quality:  212.50            Length:      85
              Ratio:  3.269               Gaps:       4
 Percent Similarity:  76.471   Percent Identity: 55.294

alignment block:
 US-09-126-945B-2 x BE553236  ..

 Align seq 1/1  to: BE553236  from: 1  to: 622

     249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 295
         :::::|||||||||||||||||     ||||||     |||   |||
     160 CTTCAGCTCTGGCAGTTTTGGTAGCTCTTCTGGATGACCCT...TCAAA 206

     265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLysI 281
          :::::|||||   |||   :::      :::|||:::   ||||||:
     207 TTCTCATTTTATTGCCTGGACTGGA.....CGAGGCATGGAATTTAAAC 250

     281 IeGluAspSerAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArg 297
         ::   :::       :::||||||||||   ||||||||||:::||||||||
     251 TGATTGAGCCTGAAGAGGTGGCCCGGCGTTGGGGCATTCAGAAGAACAGG 300

     298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 344
```

```
     |:||||||||||||||||||||||||||||||||||:::|||   ||||||::
 301 GCGGCGATGAACTATGACAAACTTAGTCGTTCTCTCGGCTATTATTATGA 350

 314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
     :|||||||||::::::|||          :::::::|||   ||||||:::|
 351 GAAGGGAATCATGCAAAAGGTG...GCTGGAGAAAGATACGTGTACAAAT 397

 331 heVal 332
     ||||
 398 TTGTG 402

seq_name: gb_est71:BE200487

seq_documentation_block:
LOCUS       BE200487      640 bp    mRNA            EST       26-JUN-2000
DEFINITION  uq63t08.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1547079 3' similar to TR:O70273 O70273 EIS HOMEOBOX FACTOR
            ;, mRNA sequence.
ACCESSION   BE200487
VERSION     BE200487.1  GI:8712656
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     other_ESTs: uq63t08.y1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:952427
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1..640
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1547079"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      133 a    155 c    135 g    216 t      1 others
ORIGIN


alignment_scores:
           Quality:  212.00            Length:     163
             Ratio:  2.186               Gaps:       4
 Percent Similarity:  59.509   Percent Identity:  30.675

alignment_block:
  US-09-126-945B-2 x BE200487/rc

  Align seq 1/1  to reverse of: BE200487  from: 1  to: 640

 169 ProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAl 185
      |||      |||:::::   :::   ||||||:::|||    ::::|||
 616 CCCAGCTATGGTAGTACAGTAGATATGTTGGCCAGTAAGACTATCTGCCG 567

 185 aMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValI 202
     :::   ::::::              :::|||
 566 GGCTCAGATCTCCATG...........ACACCT................ 545

 202 euHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArg 218
        :::|||||||   :::    :::|||            |||:::
 544 ..TCCAGTCACCTTCCAGTTGCAGAGTCACGTGATATGAAATAGGAGCAA 497

 219 ThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerIl 235
          |||          |||
 496 GACCACCCTGTAAAGTCCCAC............................ 476

 235 pThrAspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuT 252
                 ::::::   :::    :::         ||||||
 475 ................ACCANAAAGCACAACCCAAGAGGCACTCACTTAT 442

 252 rpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPhe 268
      |||:::|||::::::::::::|||||||::::|||    :::          :::
 441 GGGAGTTCATCCGAGACATTCTCTTGAGCCCAGACAAGAACCCAGGACTG 392

 269 IleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAl 285
      |||:::|||    ::::::::::|||||||||||:::::         |||
 391 ATCAAATGGGAAGACCGTTCGGAAGGCATCTTCAGGTTCCTGAAGTCAGA 342

 285 aGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnT 302
         ||||||:::||||||||      ::::|||||||:::    :::||:::::
 341 AGCTGTGGCTCAGCTGTGGGGAAAAAGAAAAATAACAGTAGCATGACAT 292

 302 yrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIle 318
      ||:::||||||||||||||:::::|||     ||||||||||:::    |||:::
 291 ACGAGAAGCTCAGCCGGGCTATGACATATTACTACAAACGAGAAATCCTG 242

 319 ArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
      ::::::    |||    :::::||||||||||||||:::|||
 241 GAACGTGTGGAT...GGACGACGGCTGGTCTATAAGTTT 206

seq_name: gb_est47:AW464218

seq_documentation_block:
LOCUS       AW464218      493 bp    mRNA            EST       24-FEB-2000
DEFINITION  BP230015A10H1 Soares normalized bovine placenta Bos taurus cDNA
            clone BP230015A10H1 5', mRNA sequence.
ACCESSION   AW464218
VERSION     AW464218.1  GI:7034386
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
            ,J.H.
  TITLE     Bovine ESTs
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Lewin, H. A.
            W. M. Keck Center for Comparative and Functional Genomics
            University of Illinois at Urbana-Champaign
            340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
            61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5617
            Email: h-lewin@uiuc.edu
            Funding for cattle EST sequencing was provided by the USDA National
            Research Initiative, Animal Genome Resource Grant AG 99 3520 #534
            to H. A. Lewin and J. E. Womack. Base calling/quality scores: PHRED
            from Washington University Genome Center. Vector Trimming:
            Cross_match from Washington University Genome Center PHRAP suite.
            Sequences submitted are vector free and at least 200 bp in length.
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Insert Length: 493   Std Error: 0.00
            Plate: BP230015A10  row: H  column: 1
```

```
            Seq primer: AGCGGATAACAATTTCACACAGA
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism "Rob Harris"
                     /db_xref "taxon:9913"
                     /clone "BB280015A10BI"
                     /clone_lib "Soares normalized bovine placenta"
                     /sex "female"
                     /lab_host "DH10B"
                     /note "Organ: placenta; Vector: pT7T3Pac; Site 1: EcoRI;
                     Site 2: NotI; the cDNA library was contributed by the
                     Soares laboratory and it was constructed and normalized
                     as described by Bonaldo, M.F., Lennon, G. and Soares,
                     M.B. (1996), Genome Research 6(9): 791 806. "
BASE COUNT       148 a    113 c    118 g    119 t
ORIGIN


alignment scores:
            Quality:  211.50             Length:      84
              Ratio:  3.065                Gaps:       4
 Percent Similarity:  82.143   Percent Identity: 54.571

alignment block:
 US-09-126-945B-2 x AW464218   ..

 Align seq 1/1  5... AW464218  from: 2  to... 493

      249 HisHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
          |||  |||||||||||||||  :::||||||  :::|||::::::
        4 ATCACCCTGTGGCAGTTCCTCCTTCAGCTCCTAAAGGAGCCTCAGAAC.. 51

      265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleG 282
            :::::|||   |||  :::::::::::|||  |||||:::
       52 .GACCACATGATCTGCTGG...ACCTCTAATAATGGGGAGTTCAAGCTTT 97

      282 lnAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgPro 298
          ::::::  :::||||||||||||||||||||||||||||||||:::|||
       98 TGCAGGCAGAAGAGGTGGCTGGACTCTGGGGCATTCGAAAGAACAAGCCT 147

      299 AlaMetAspTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysTy 315
          ||||||||||||||||||||||:::::|||  ||||||  ||
      148 AATATGAATTATGACAAACTCAGCCGACGCCTCAGATACTACTATGTGAA 197

      315 sGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnTheV 332
          |:::|||||||:::|||  :::  :::|||:::::|||||||:::||||
      198 GAATATCATTAAAAAAGTGAAT...GGTCAGAAGTTTGTGTACAACGTTG 244

      332 al 332
          ||
      245 TC 246

seq name: qb_est26:AI877751

seq documentation block:
LOCUS        AI877751        718 bp    mRNA            EST       21 JUL 1999
DEFINITION   to51bl1.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
             TR:o57586 o57586 ETS DOMAIN TRANSCRIPTION FACTOR PEA3. ;, mRNA
             sequence.
ACCESSION    AI877751
VERSION      AI877751.1  GI:5551800
KEYWORDS     EST.
SOURCE       zebrafish.
  ORGANISM   Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE    1  (bases 1 to 718)
  AUTHORS    Clark,M., Johnson,S.L., Lehrach,H., Lee,K., Li,F., Marra,M., Eddy
             ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
             ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
             Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,K., Ritter,E.,
             Kohn,S., Shin,I., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
             and Wilson,R.
  TITLE      WashU Zebrafish EST Project 1998
  JOURNAL    Unpublished (1998)
COMMENT      other ESTs: to51bl1.x1
             Contact: Stephen L. Johnson
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: zbrafish@watson.wustl.edu
             cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
             Matthew Clark. DNA Sequencing by: Washington University Genome
             Sequencing Center Clone distribution: Genome Systems, St. Louis,
             Missouri (web address: www.genomesystems.com) (email contact:
             info@genomesystems.com) and Research Genetics, Huntsville, Alabama
             (web address: www.resgen.com) (email contact: info@resgen.com) and
             RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
             www.rzpd.de)
             Possible reversed clone; similarity on wrong strand
             Seq primer: T3 EI from Amersham
             High quality sequence stop: 478.
FEATURES             Location/Qualifiers
     source          1. .718
                     /organism "Danio rerio"
                     /db_xref "taxon:7955"
                     /clone_lib "Zebrafish WashU MPIMG EST"
                     /sex "mixed"
                     /tissue_type "26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host "XL1 blue MRF"
                     /note "Vector: pSPORII; Site 1: NotI; Site 2: SalI; 1st
                     strand cDNA was primed with a Not I  oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT3']
                     double stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORII vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab, ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (24 hpf), 48hf liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT       173 a    194 c    191 g    158 t      2 others
ORIGIN


alignment scores:
            Quality:  211.00             Length:     332
              Ratio:  1.407                Gaps:      14
 Percent Similarity:  45.181   Percent Identity: 25.602

alignment block:
 US-09-126-945B-2 x AI877751   ..

 Align seq 1/1  to: AI877751  from: 1  to: 718

        3 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
          |||||||:::     ::::::   :::   |||:::|||
        8 TCCGCGGACCGCTGGGTGCAGCGCATGCATCCTGCCCAT.......... 46

       19 oProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaV 36
                                          ::::::::|||:::
       47 ..............................GCCAGCGGAGGCGGACGAG 65

       36 alGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu 52
            |||   :::|||:::      :::|||     |||     |||:::
```

```
 66 GCGGCTACCATCGGCAACACTCTGACCCCTGCTTGCCTTATCCTCCTCAG 115

 54 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl 69
    |||          ::|||      :::|||::::||   ||||||   ::
116 CAG.........ACCTTTAAGCAGGAGTACATGGACCCTCTGTATGACCG 156

 69 uAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluP 86
    :      ||||||           |||   :::   ::::::    |
157 A.........GCGGCCCACATTAACGGACCACAGCCTCAGAGGTTTC 191

 86 roProGluGlu......ProGluGlnCysProValIleAspSerGlnAla 100
    |||||   :::      ::|||   ||||::      :::
196 CTCCTGCCCATATGATGGTCAAACAGGAGCCCACCGACTACACTTACGAA 244

101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHi 117
    |||                 |||||||||       ::::::   :::::
245 CCTGAT.............GTGGCTGGCTGCCCATCCATGTACTATC 279

117 sSerLeuGlnGlnValGlnSerMetValValGlyGluValLeuLysAspI 134
    ||::
280 CAAI.......................................... 283

134 leGluThrAlaCysLysGlyGlnLeuAspIleThrAlaAspIleMetAspTrp 155
                                    :::     :::
284 ..................................GAAGGCTACTCCAAC 298

151 SerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGluTyrAr 167
              |||   :::   :::::::::      |||::::::::   ||
299 CCACAGCACAACAGTGAAGGCTACATGTTT.....GAAAATGATTCCCG 342

167 qLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluGlnT 184
    ||::   |||
343 TGTTGTGCCA..................................... 352

184 ysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAsp 203
                 |||::::|||::                  |||::
353 ............GAGAAATTGAA................GGTGAC 376

201 ValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGl 217
    |||   :::        ::::::::::::::::|||      :::::
371 GTGAAACAAGAAGGGGCACTGTGTTTCGTGAAGCTGCCCCCTATCAGCG 420

217 uArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGlnS 234
    :|||   |||
421 TCGTGGCTCA..................................... 430

234 erTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHis 253
                                                 ::::::
431 ..............................................CTTCAA 436

251 LeuTrpGluPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyAr 267
    ::|||||||||||||      ||||||      |||   :::   :::::
437 CTATGGCAGTTCTTGGTTGGCCTTCTTGATGACCCC...AGCAATGCCCA 486

267 qPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLysIleGluA 283
    :|||||   |||   :::      ::::|||::   ||||||:::   :
484 CTTCATCGCATGGACGGGC......CGTGGCATGGAAATTCAAACTCATTG 527

283 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299
    ::      :::||||||||||||||||||||:::::|||||||||||||||
528 AGCCAGAGCAGGTGGCAAGACTCTGGGGGATGCAGAAGAACCGCCCAGCC 577

300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
    ||||||||||||||||||||||||::::||   ||||||::||||||
578 ATGAACTATGACAAACTGAGTCGCTCTTTGCGTTACTATTATGAGAAGGG 627

316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
    |||:::   |||         :::         ||||||||||||
628 AATTATGCCAAAGGTG...GCTGGGGCAGCGTATGTTTACCAATTT 670

seq_name: gb_est89:BF571534

seq_documentation_block:
```

```
LOCUS       BF571534       853 bp    mRNA            EST       12 DEC 2000
DEFINITION  602078199F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4252616 5',
            mRNA sequence.
ACCESSION   BF571534
VERSION     BF571534.1  GI:11645246
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   NIH MGC Project
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ARS/Beltsville/BIP
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAMHU77  row: n  column: 09
            High quality sequence start: 3
            High quality sequence stop: 468.
FEATURES             Location/Qualifiers
     source          1..853
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4252616"
                     /clone_lib="NIH_MGC_62"
                     /tissue_type="melanotic melanoma, High MDR"
                     /lab_host="DH10B (T1 phage resistant)"
                     /vector="pCMV-SPORT6 (CloneTech)"  Site_1:
                     SfiI (ggccgctctagcc)  Site_2:  SfiI (ggccattatggcc)
                     Double-stranded cDNA was prepared from cell line kNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-d(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T).  Average
                     insert size 1.75 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Cloutech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      195 a    226 c    249 g    183 t
ORIGIN


alignment_scores:
           Quality:  210.50              Length:     83
             Ratio:   5.289                Gaps:      4
 Percent Similarity:  77.108   Percent Identity: 55.422

alignment_block:
 US-09-126-945B-2 x BF571534   ..

 Align seq 1/1  to: BF571534  from: 1  to: 853

     251 LeuTrpGluPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyAr 267
         ::|||||||||||||      ||||||      |||   :::   :::::
      80 GTGTGGCAATTCTTGGTTGGCCTTCTTGATGGATGACCCA...ACAAATGCCCA 126

     267 qPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLysIleGluA 283
         :|||||   |||   :::      ::::|||::   ||||||:::   :
     127 TTTCATTGCCTGGACGGGC......CGGGGAATGGAGTTCAAGCTCATTG 170

     283 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299
         ::      :::||||||||||||||||||||:::::|||||||||||||||
     171 AGCCTGAGGAGGTCGCCAGGCTCTGGGGCATCCAGAAGAACCGGCCAGCC 220

     300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
         ||||||||||||||||||||||||||::::|||   ||||||::||||||
```

```
      221 ATGAATTACGACAAGCTGAGCCGCTCGCTCCGATACTATTATGAGAAACGC 270

      316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal 332
          ||||:::::::|||          :::::::|||   |||||||::|||||||
      271 CATCATGCAGAAGGTG...GCTGGTGAGCGTTACGTGTACAACTTTGTG 316

seq name: qb est84:BF144998

seq documentation block:
LOCUS       BF144998        733 bp    mRNA            EST       24 OCT 2000
DEFINITION  601790295F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4021079 5',
            mRNA sequence.
ACCESSION   BF144998
VERSION     BF144998.1  GI:10984038
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 733)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            cDNA Sequencing by:Incyte Genomics, Inc.
            Clone distribution: NCI CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9279  row: n  column: 24
            High quality sequence stop: 691.
FEATURES             Location/Qualifiers
     source          1..733
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4021079"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV LTR enhancer; Cloned unidirectionally.  Primer: oligo
                     dT.  Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      188 a    186 c    187 g    172 t
ORIGIN


alignment scores:
            Quality:  210.00            Length:    196
              Ratio:  1.707               Gaps:      4
 Percent Similarity:  62.755   Percent Identity:  42.143

alignment block:
 03 09 126 945B 2 x BF144998  ..

 Align seq 1/1  to: BF144998  from: 1  to: 733

      145 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
          :::|||||||||||          ::::::  |||      ||||:
      152 CAGACAGCCTGTGATTCCTACTGGACATCAGTGCACCCTGAATACTGGAC 201

      151 rProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 199
          :    :::|||   :::|||||   :::         |||||||:::|
      202 CAAGCGCCACGTCTGGGAATGGCTCCAATTCTGCTGCGACCACTACAAGC 251

      168 euProProMetGlyLysAlaPh......GluGluLeuAlaGlyLysGlu 162
          ||              ::|||           ::::::::|||   :::
      252 TTGATGCCAACTGCATCTCCTTCTGTCACTTCAACATCAGCGGCCTGCAG 301

      183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199
          |||||||:::|||:::::|||:::|||   :::   :::   :::   ||
      302 CTCTGCAGCATGACGCAGCAGCAGTTCATTGAGGCAGCCGGCATCTGTGG 351

      199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMet. 215
          |:::   ||||::      |||:::      :::|||   ::::::
      352 GGACTACCTGTACTTCATTCTCCAGAACATTCGCCTCGCAAGGTTACTCCT 401

      216 ..........LysGluArgThrSerProGlyAlaIleHisTyr..... 226
                     ||||:::|||::   |||         |||
      402 TTTTCAATGATGCTGAAGAAACTAAGACTGGCATCAAAGACTATGCTGAT 451

      227 ......CysAlaSerThrSerGluGluSerTrpThrAspSerGluValAs 241
                ||||   :::|||||||      :::      :::::::::
      452 TCCAGTTGCTTGAAAACAACT.....GGCATCAAGAGTCAACACTGTCA 495

      241 pSerSerCysSerGlyGlnProIleHisLeuTrpGluPheLeuLysGluL 258
          |||      |||   |||       ||||||||||::|||::::::::::|
      496 CAGCCGAACAAGCCTCCAAAGTTCTCACCTGTGGGAATTTGTCAGAGACT 545

      258 euLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 274
          |||||   :::|||::::::          :::::::::|||   ::::::
      546 TGCTGTGCTCCCCTGAAGACAACTGTGGCATCCTGGAATGGGAAGACAGG 594

      275 GluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTr 291
          |||:::|||||||||:::::::      |||      :::|||:::::||
      595 GAGCAGGGCATTTTCCGAGTGCTTAAGTCAGAAGCCCTGGCAAAGATGTG 644

      291 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 308
          ||||   ||||||::::::          |||:::|||:::|||||||||||
      645 GGGACAACGAAGAAGAATGACAGGATGACGTACGAGAAGTTGAGCGCAG 692

      308 erIleArgGlnTyrTyrLysLysGlyIleIleArgLys 320
          :::|||   |||||||||||||||||||||:::::::::
      693 CTTGAGATACTATTATAAAAAGGGGATTCTGCAGCGC 730

seq name: qb est103:C46450

seq documentation block:
LOCUS       C46450          369 bp    mRNA            EST       18 OCT 1999
DEFINITION  yk418d10.5 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
            Caenorhabditis elegans cDNA clone yk418d10 5', mRNA sequence.
ACCESSION   C46450
VERSION     C46450.1  GI:2382703
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
            ,M., Miyata,A. and Nishiwaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81 559 81 6854
            Fax: 81 559 81 6855
            Email: ykohara@lab.nig.ac.jp
FEATURES             Location/Qualifiers
     source          1..369
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk418d10"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT      116 a     88 c     82 g     83 t
```

```
ORIGIN


alignment_scores:
            Quality:  209.50              Length:      97
              Ratio:  2.993                Gaps:       4
 Percent Similarity:  72.165   Percent Identity: 49.485

alignment_block:
 US-09-126-945B-2 x C46450   ..

 Align seq 1/1  to: C46450  from: 1  to: 369

      236 ThrAspSerGlnValAspSerSerCysSerGlyGlnIleHisLeuIl 275
          |:|:::|:|:::|||||||||                 |||   |||||
       44 ACCTCATCATCCGTCGACTCAATC..............AIAACCCTGIG 78

      252 pGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheI 299
          ||||||||||   ||||||||||   :::   ::::::   |||   :::|
       79 GCAATTCCTICTAGAACTACTGCAACAAGACCAAAAT...GGTGATATAA 125

      269 leArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAla 285
          |||:|||   :::           |||   |||::::   ||:::
      126 TCGAATGGACACGCGGAACGGACGGCGAATTCCGACTGATTGATGCAGAA 175

      286 GlnValIleArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTy 302
          ||||||||   |:|||   ||||||   :::|||   ||||||||
      176 GCCGTGGCGAGAAAGTGGGGACAACGGAAGGCGAAACCGCATATGAATTA 225

      302 rAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleA 319
          |||:|||||||||||||:::::|||   |||||||:::|||:::|:|||
      226 TGATAAACTGTCGCAGAGCGTTACGATATTATTATGCAGAAGAATATTATA 275

      319 rgLysProAspIleSerGlnArgLeuValTyrGlnPheVal 352
          :::||      |||:::::::::::::::::::::::|||
      276 AGAAGGTG...ATCGGCAAAAAGTTCGTATATCGCTTTGTA 313

Seq_name: qb_est94:BF896156

Seq_documentation_block:
LOCUS       BF896156        472 bp    mRNA            EST       18-JAN-2001
DEFINITION  CM2-MT0157-231100-545-e06-MT0157 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF896156
VERSION     BF896156.1  GI:12287515
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brumstein,A., deOliveira,P.S., Bucher,P., Jothirenedi,C.V., o'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2701922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/cgi-bin/db/pathcall?file=MT/CLS_MT-MT0157-
            231100-545-e06&f3 2000-11-24&f4 -1)
            Seq primer: puc 18 forward
            High quality sequence start: 22
            High quality sequence stop:  343.
FEATURES             Location/Qualifiers
     source          1..377
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib: "MT0157"
                     /dev_stage="Adult"
                     /note "Organ: marrow; Vector: puc18; Site 1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 09
                     /716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT        71 a    103 c     99 g    104 t
ORIGIN


alignment_scores:
            Quality:  209.50              Length:      79
              Ratio:  3.379                Gaps:       1
 Percent Similarity:  78.481   Percent Identity: 48.101

alignment_block:
 US-09-126-945B-2 x BF896156/rev   ..

 Align seq 1/1  to reverse of: BF896156  from: 1  to: 377

      253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl 269
          :::|||||||     :::|||||||     ::::::     :::::::||
      377 CAATTCCTCCTGGCTTTTCTGCCTAGACATATACACCTGTCCCAACTACAT 328

      269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaG 286
          |::::||   ::::::|||||||||||||||||:::   ||||||
      327 CAAGTGGACCCAGCGAGAGAAAGGTATCTTCAAACTGGTGGACTCCATAG 278

      286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
            |||:::::|||||||||   :::|||||||:::|||   |||||||||
      277 CTGTGTCCAAGCGTGTGGGGAAGCAGAAAAACAAGCCGACATGAACTAT 228

      303 AspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 319
          :::   ::::::|||:::::|||   |||||||:::::|||||:::
      227 CAGACAATCGAGCGTCGCCACTAAGATACTACTACCAAAGAGGCATACTGGC 178

      319 gLysProAspIleSerGlnArgLeuValTyrGlnPhe 351
          |||   :::   :::|||||||||||||||||||||
      177 CAAAGTGGAA...GGGCAGAGGCTGGTGTACCAGTT 144

Seq_name: qb_est77:BE626845

Seq_documentation_block:
LOCUS       BE626845        403 bp    mRNA            EST       24-AUG-2000
DEFINITION  ut88c01.xl Soares mammary_gland NbMG Mus musculus cDNA clone
            IMAGE:3469504 3' similar to TR:O70273 O70273 EIS HOMOLOGOUS FACTOR
            ;, mRNA sequence.
ACCESSION   BE626845
VERSION     BE626845.1  GI:9907265
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   NCI-CGAP Mouse Genome Anatomy Group.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: ut88c01.y1
            Contact: Robert Strausberg, Ph.D.
            Email: cgaphs-email@nih.gov
            This clone is available royalty free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
```

```
            MGI:1079108
            Trace considered overall poor quality
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1..404
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3469504"
                     /clone_lib="Soares mammary gland NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I  oligo(dT) primer. Double stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       85 a     103 c      86 g     129 t
ORIGIN


alignment scores:
            Quality:  209.50              Length:      82
              Ratio:  3.081                 Gaps:       1
 Percent Similarity:  82.927   Percent Identity: 45.122

alignment block:
 DS-09-126-9458-2 x BE626845_rc  ..

 Align seq 1/1  to reverse of BE626845  from 1  to 403

      250 HisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGl 266
          ||||||||||:::|||:::::::::::::|||||||::|||   :::   ::
      277 CACTTATGGGAGTTCATTCGAGACATTCTCTTGAGCCCAGACAAGAACTC 228

      266 yArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluA 283
          :   :::|||:::|||   :::::::::::::|||||||:::::::
      227 AGGACTGATCAAATGGGAAGACCGTTCGGAAGGCATCTTCAGTTCCTGA 178

      284 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299
            |||       |||||:::|||||||||   :::|||||||:::   :::
      177 ACTCAGAACCTCGTGGCTCAGCTGTGGGGAAAAAGAAAAATAACAGTAGC 128

      300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
          |||:::|||:::|||||||||||||:::::::|||   ||||||||||:::
      127 ATGACATACGAGAAGCTCAGCCGGCTATGAGATATTACTACAAACGAGA 78

      316 yIleIleArgLysProAspIleSerGluArgLeuValTyrGlnPhe 331
           |||:::::::::   |||   :::::::|||||||||||:::|||
       77 AATCCTGGAACGTGTGGAT...GGACGACGGCTGGTCTATAAGTTT 35

seq_name: qb_est26:AI876285

seq_documentation_block:
LOCUS        AI876285          422 bp    mRNA            EST       21-JUL-1999
DEFINITION   ak73d07.y1 Schiller mouse AIT20 Mus musculus cDNA 5', clo
             IMAGE:1974637 5' similar to gb:L10426 Mus musculus ets related
             protein 81 (MOUSE);, mRNA sequence.
ACCESSION    AI876285
VERSION      AI876285.1  GI:5550434
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 422)
  AUTHORS    Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
             Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
             ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
             ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
             Waterston,R. and Wilson,R.
  TITLE      The WashU-NCI Mouse EST Project 1999
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Marra M/WashU-NCI Mouse EST Project 1999
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouse@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info-image.llnl.gov) for further information.
             MGI:991477
             Seq primer:  40RP from Gibco
             High quality sequence stop: 384.
FEATURES             Location/Qualifiers
     source          1..422
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1974637"
                     /clone_lib="Schiller mouse AIT20"
                     /cell_line="pituitary cell line"
                     /lab_host="SOLR"
                     /note="organ: pituitary; Vector: pBluescript SK
                     (Stratagene); Site 1: EcoRI; Site 2: XhoI; Double stranded
                     cDNA was prepared from cell line AIT 20 using primer
                     5' GAGAGAGAGAGAGAGAGAGAAACTAGTCTCGAGT(18) 3'.  An EcoRI
                     adaptor was used on the 5' end of the cDNA as follows:
                     5' AATTCGGCACGAG 3'.  The library was size selected and
                     went through one round of amplification.  Average insert
                     size is 1.7 kb, with a range from 0.4 to 4 kb.  This library
                     was constructed by Dr. Martin Schiller (Johns Hopkins
                     University). "
BASE COUNT       101 a     113 c     116 g      92 t
ORIGIN


alignment scores:
            Quality:  209.50              Length:      85
              Ratio:  3.223                 Gaps:       4
 Percent Similarity:  76.471   Percent Identity: 54.118

alignment block:
 DS-09-126-9458-2 x AI876285  ..

 Align seq 1/1  to AI876285  from 1  to 422

      249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
          ::::::|||||||||||||||||       ||||||      |||   :::
      122 CTGCAGCTGTGGCAGTTCCTTGTCACCCTCCTGGATGACCCA...GCCAA 168

      265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLysI 281
          ::::::::::::|   |||   :::      :::|||:::   ||||||
      169 TGCCCACTTCATCGGCCTGCACTGGC.....CGAGGCATGGAATTTAAGC 212

      281 leGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
          ::   :::      :::||||||||   ||||||||||:::||||||||
      213 ICATAGAACCGGAAGAGGTTGCTCGCCGTTGGGGTATCCAGAAGAATCGG 262

      298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyrTyrLy 314
          ||||||||||||||||||||||||||||||||:::|||   ||||||::
      263 CCAGCCATGAACTATGACAAGCTGAGCGCTCTCATCCGCTATTACTATGA 312

      314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
          :||||||||:::::::||          :::::|||   ||||||:::|
      313 AAAGGGCATCATCAGAAAGATG...GCTGGGGAACGCTACGTCTACTATAA 359

      331 heVal 332
          |||||
      360 TGTG 363

seq_name: qb_est106:N29134

seq_documentation_block:
```

```
LOCUS       N29134        500 bp    mRNA            EST       04 JAN 1996
DEFINITION  yx47t05.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone
            IMAGE:264897 5' similar to SP:SW:ERM_HUMAN P41161 ETS RELATED
            PROTEIN ERM. [1] ;, mRNA sequence.
ACCESSION   N29134
VERSION     N29134.1  GI:1147370
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 432
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1423   Std Error: 0.00
            Seq primer: T7
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1..500
                     /organism "Homo sapiens"
                     /db_xref="GDB:3874539"
                     /db_xref="taxon:9606"
                     /clone "IMAGE:264897"
                     /clone_lib "Soares melanocyte 2NbHM"
                     /sex "Male"
                     /tissue_type="melanocyte"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note "Vector: pT7T3D (Pharmacia) with a modified
                     polylinker. Site_1: Not I, Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo (dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo. RNA from normal foreskin melanocytes
                     (FS374) was kindly provided by Dr. Anthony P. Albino."
BASE COUNT      115 a    145 c    132 g    107 t      1 others
ORIGIN


alignment_scores:
           Quality:  209.50              Length:     85
             Ratio:  3.223                 Gaps:      4
 Percent Similarity:  76.471   Percent Identity: 54.118

alignment_block:
 US-09-126-945B-2 x N29134   ..

 Align seq 1/1  to: N29134  from: 1  to: 500

     249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerGly 265
         ::::::|||||||||||||||       ||||||      |||    :::
     114 CTTCAGCTGTGGCAGTTCCTGGTCACCCTCTTGATGACCCA...GCCAA 160

     265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLysT 281
          ::::::||||||   |||   :::       :::|||:::   ||||||:
     161 TGCCCACTTCATTGCCTGGACAGGT......CGAGGCATGGAGTTCAAGC 204

     281 leGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
         ::    :::        :::|||||||||    ||||||||:::||||||||
     205 TGATAGAACCGGAAGAGGTTGCTCGGCGCTGGGGCATCCAGAAGAACCGG 254

     298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 314
         |||||||||||||||||||||||||||||||||||:::|||    ||||||::
     255 CCAGCCATGAACTATGACAAGCTGAGCCGCTCTCTCCGCTATTACTATGA 304

     314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
         :||||||||:::::::|||              :::::::|||    |||||||:::|
     305 AAAGGGCATCATGCAGAAGGTG...GCTGGAGAGCGATACGTCTACAAAT 351

     331 heVal 332
         |||||
     352 TTGTC 356

seq_name: gb_est82:BF021513

seq_documentation_block:
LOCUS       BF021513        516 bp    mRNA            EST       29 DEC 2000
DEFINITION  uy48a11.y1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:3662780 5'
            similar to SW:ERM_HUMAN P41161 ETS-RELATED PROTEIN ERM ;, mRNA
            sequence.
ACCESSION   BF021513
VERSION     BF021513.1  GI:10752845
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1423548
            Seq primer:  40RP from Gibco
            High quality sequence stop: 415.
FEATURES             Location/Qualifiers
     source          1..516
                     /organism "Mus musculus"
                     /strain "CZECH II"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:3662780"
                     /clone_lib "NCI_CGAP_Lu30"
                     /tissue_type "tumor, metastatic to mammary"
                     /lab_host "DH10B"
                     /note "Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT 1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally. Primer: oligo
                     dT.  Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      119 a    155 c    137 g    105 t
ORIGIN


alignment_scores:
           Quality:  209.50              Length:     85
             Ratio:  3.223                 Gaps:      4
 Percent Similarity:  76.471   Percent Identity: 54.118

alignment_block:
 US-09-126-945B-2 x BF021513   ..
```

```
Align Seq 1/1  to: BE021513  from: 1  to: 516

     249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
         :::::!!!!!!!!!!!!!!!!      !!!!!!      !!!   :::
     140 CTGCAGCTGTGGCAGTTCCTCCTCCAGCCTCCTGGATGACCCA...GCCAA 176

     265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLys1 281
          :::::!!!!!!   !!!   :::      :::!!!:::   !!!!!!:
     177 TGCCCACTTCATCGCCTGGACTGGC.....CGAGGCATGGAATTTAAGC 229

     281 IeGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
         ::   :::      :::!!!!!!!!!   !!!!!!!!!:::!!!!!!!!
     221 TCATAGAACCGGAAGAGGTTGCTCGCCGTGGGGTATCCAGAAGAACCGG 270

     298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 314
         !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!:::!!!   !!!!!!:
     271 CCAGCCATGAACTATGACAAGCTGAGCCGCTCTCTCCGCTATTACTATGA 320

     314 sLysGlyIleIleArgLysProAspIleSerGluArgLeuValTyrGlnP 331
         :!!!!!!!!!!:::::!!!          :::::!!!   !!!!!!:::!
     321 AAAGGGCATCATGCAGAAGGTG...GCTGGGGAACGCTACGTCTACAAAT 367

     331 hoVal 332
         !!!!!
     368 TTGTG 372

Seq_name: qb_est41:AU140178

Seq_documentation_block:
LocUS        AU140178       863 bp     mRNA             EST       25 OCT 2000
DEFINITION   AU140178 PLACE2 Homo sapiens cDNA clone PLACE2000081 5', mRNA
             sequence.
ACCESSION    AU140178
VERSION      AU140178.1  GI:11001699
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 863)
  AUTHORS    Ota,T., Suzuki,Y., Saito,K., Ishii,S., Yamamoto,J., Sugiyama,F.,
             Nishikawa,F., Nakamura,Y., Sugano,S., Masuho,Y. and Isogai,T.
  TITLE      HRI human cDNA project (Ota,T., Suzuki,Y., Saito,K., Ishii,S.,
             Yamamoto,J., Sugiyama,F., Nishikawa,F., Nakamura,Y., Sugano,S.,
             Masuho,Y., Isogai,T.)
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Takao Isogai
             Genomics Laboratory
             Helix Research Institute
             1532 3 Yana, Kisarazu, Chiba 292 0812, Japan
             Tel: 81 438 52 3951
             Fax: 81 438 52 3952
             Email: genomics@hri.co.jp
             HRI human cDNA project; 5'  & 3' end one pass sequencing: Helix
             Research Institute; cDNA library construction: Department of
             Virology, Institute of Medical Science, University of Tokyo, and
             Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..863
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "PLACE2000081"
                     /clone_lib "PLACE2"
                     /tissue_type "placenta"
                     /note "Vector: pMEI8SFL4"
BASE COUNT       211 a    229 c    225 g    191 t      7 others
ORIGIN


alignment scores:
            Quality:  209.50            Length:     85
              Ratio:  3.223               Gaps:      1
  Percent Similarity:  76.471   Percent Identity:  54.118

alignment_block:
 US-09-126-945B-2 x AU140178  ..

 Align Seq 1/1  to: AU140178  from: 1  to: 863

     249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
         :::::!!!!!!!!!!!!!!!!      !!!!!!      !!!   :::
     319 CTGCAGCTGTGGCAGTTCCTGGTCACCCTCCTGGATGACCCA...GCCAA 365

     265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLys1 281
          :::::!!!!!!   !!!   :::      :::!!!:::   !!!!!!:
     366 TGCCCACTTCATCGCCTGGACTGGC.....CGAGGCATGGAATTTAAGC 409

     281 IeGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
         ::   :::      :::!!!!!!!!!   !!!!!!!!!:::!!!!!!!!
     410 TGATAGAACCGGAAGAGGTTGCTCGCCGTGGGGCATCCAGAAGAACCGG 459

     298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 314
         !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!:::!!!   !!!!!!:
     460 CCAGCCATGAACTATGACAAGCTGAGCCGCTCTCTCCGCTATTACTATGA 509

     314 sLysGlyIleIleArgLysProAspIleSerGluArgLeuValTyrGlnP 331
         :!!!!!!!!!!:::::!!!          :::::!!!   !!!!!!:::!
     510 AAAGGGCATCATGCAGAAGGTG...GCTGGGAGAGCGATACGTCTACAAAT 556

     331 hoVal 332
         !!!!!
     557 TTGTC 561

Seq_name: qb_est52:AW853070

Seq_documentation_block:
LocUS        AW853070       511 bp     mRNA             EST       19 MAY 2000
DEFINITION   BCI-CI3219 020408 026 a05 CI0219 Homo sapiens cDNA, mRNA sequence.
ACCESSION    AW853070
VERSION      AW853070.1  GI:7948587
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 511)
  AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brumstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE      Shotgun Sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491 3496 (2000)
  MEDLINE    20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509 010, Sao Paulo SP,
             Brazil
             Tel: +55 11 2704922
             Fax: +55 11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (URL: /www-berr.cr.bcm.tmc.edu/cgi-bin/script_sptbtmiz.pl?1 &t2 BCI-CI0219-020
             408-026-a05&t3 2000-05-03&t4 1)
             Seq primer: puc 18 forward
             High quality sequence start: 15
             High quality sequence stop: 510.
FEATURES             Location/Qualifiers
     source          1..511
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
```

```
                     /clone_lib "CT0249"
                     /dev_stage "Adult"
                     /note "Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector.  Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       177 a    116 c    120 g     98 t
ORIGIN


alignment_scores:
            Quality:  209.00              Length:     128
              Ratio:  2.549                 Gaps:       2
 Percent Similarity:  64.062   Percent Identity:  42.812

alignment_block:
 US-09-126-945B-2 x AW853079    ..

 Align seq 1/1  to: AW853079  from: 1  to: 511

      204 AlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgIhrS 220
          :::|||||   :::   :::|||              |||||||:::
      194 AGTCACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCC 153

      220 rProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrA 267
           |||:::      |||
      154 CCCTGCCAAGTGCCAC.................................. 169

      267 spSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGln 253
                    ::::::   :::   :::      ||||||||||:::
      170 ...........ACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGGAA 208

      254 PheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleAr 270
          |||:::::::::::::   |||:::|||   :::      :::||||::
      209 TTCATCCGCGACATCCCCTTCAACCCAGACAAGAACCCAGGATTAATAAA 258

      270 qTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGluV 287
          :|||   :::::::::::|||:::|||:::::      |||     |
      259 ATGCGAAGACCGATCTGAGGGCGTCTTCAGTTCTTGAAATCAGAGGCAG 308

      287 alAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAsp 303
          ||||:::|||||||||   :::|||||:::   :::|||:::||||:::
      309 TGGCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTATGAA 358

      304 LysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLy 320
          ||||||||||||:::::|||   |||||||||:::   ||||:::::::::
      359 AAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATTCTGGAGGG 408

      320 sProAspIleSerGlnArgLeuValTyrGlnPhe 334
          :   |||   :::::|||||||||||||:|
      409 TGTGGAT...GGACGAAGACTGGTATATAAATTT 439

seq_name: qb_est52:AW853081

seq_documentation_block:
LOCUS        AW853081       560 bp    mRNA            EST       19-MAY-2000
DEFINITION   RC1-CT0249-030300-026-d11 CT0249 Homo sapiens cDNA, mRNA sequence.
ACCESSION    AW853081
VERSION      AW853081.1  GI:7948598
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 560)
  AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE    20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55 11 2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FANTOM/RIKEN Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.fidia-arco.it/scripts_sperber/id_p1?1_%2_RC1-CT0249-030
             300-026-d11&t3=2000-03-03&t4=1)
             Seq primer: puc 18 forward
             High quality sequence start: 8
             High quality sequence stop: 559.
FEATURES             Location/Qualifiers
     source          1..560
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib "CT0249"
                     /dev_stage "Adult"
                     /note "Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector.  Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       202 a    128 c    123 g    107 t
ORIGIN


alignment_scores:
            Quality:  209.00              Length:     128
              Ratio:  2.549                 Gaps:       2
 Percent Similarity:  64.062   Percent Identity:  42.812

alignment_block:
 US-09-126-945B-2 x AW853081    ..

 Align seq 1/1  to: AW853081  from: 1  to: 560

      204 AlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgIhrS 220
          :::|||||   :::   :::|||              |||||||:::
      132 AGTCACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCC 181

      220 rProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrA 267
           |||:::      |||
      182 CCCTGCCAAGTGCCAC.................................. 197

      267 spSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGln 253
                    ::::::   :::   :::      ||||||||||:::
      198 ...........ACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGGAA 240

      254 PheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleAr 270
          |||:::::::::::::   |||:::|||   :::      :::||||::
      241 TTCATCCGCGACATCCCCTTCAACCCAGACAAGAACCCAGGATTAATAAA 290

      270 qTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGluV 287
          :|||   :::::::::::|||:::|||:::::      |||     |
      287 ATGCGAAGACCGATCTGAGGGCGTCTTCAGTTCTTGAAATCAGAGGCAG 336

      287 alAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAsp 303
          ||||:::|||||||||   :::|||||:::   :::|||:::||||:::
      337 TGGCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTATGAA 386

      304 LysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLy 320
          ||||||||||||:::::|||   |||||||||:::   ||||:::::::::
```

```
      387 AAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATTCTGGAGCG 436

      320 sProAspIleSerGlnArgLeuValTyrGlnPhe 344
          :  |||  ::::::||||||||||||:::|||
      437 TGTGGAT...GGACGAAGACTGGTATATAAATTT 467

seq_name: qb_est52:AW853010

seq_documentation_block:
LOCUS        AW853010      619 bp    mRNA            EST       19-MAY-2000
DEFINITION   RC1-CT0249-170200-025-e01-CT0249.Homo sapiens cDNA, mRNA sequence.
ACCESSION    AW853010
VERSION      AW853010.1  GI:7948527
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 619)
  AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brumstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
  TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE    20202663
COMMENT      Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo SP,
             Brazil
             Tel: +55 11 2704922
             Fax: +55 11 2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/cgi-bin/db/CGIdb.pro?  ...
             200-025-e01&t=8 2000-02-17&14:1)
             Seq primer: puc 18 forward
             High quality sequence start: 28
             High quality sequence stop: 615.
FEATURES             Location/Qualifiers
     source          1..619
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0249"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2:
                     SmaI; A mini library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription at Clone
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       217 a    146 c    137 g    119 t
ORIGIN


alignment scores:
            Quality:  209.00              Length:    128
              Ratio:  2.549                 Gaps:      2
 Percent Similarity:  64.062   Percent Identity:  42.812

alignment block:
 US-09-126-945b-2 x AW853010  ..

 Align seq 1/1  to: AW853010  from: 1  to: 619

      204 AlaHisLeuAspIleErgLysSerAlaAlaTrpMetLysGluArgThr 239
          :::|||||||  :::  :::|||            |||||||:::
      190 AGTCACCTTCCTGTTGCAGTCACCTGATATGAAAAAGGAGCAAGACCC 239

      220 rProGlyAlaIleHisTyrCysAlaSerThrSerGlnGluSerTrpThrA 267
          |||:::   |||
      240 CCCTGCCAAGTGCCAC..................................... 255

      237 spSerGluValAspSerSerCysSerGlyGluProIleHisLeuTrpGlu 283
                  ::::::  :::  :::      ||||||||||:::
      256 ..........ACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGGAA 294

      284 PheLeuLysGlnGluLeuLysProHisSerTyrGlyArgPheIleAr 330
          |||:::::::::::::   |||:::|||  :::         :::||||:
      295 TTCATCCGCGACATCCCCTTGAACCCAGACAAGAACCCAGGATTAATAAA 344

      270 qTripLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnV 287
          :|||  ::::::::::::||:::|||:::::      |||     |
      345 ATGGGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGCCAG 394

      287 aIAlaArqLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAsp 333
          |||||:::|||||||||   :::||||||:::   :::||:::||||:
      395 TGGCTCAGCTATGGGCTAAAAACAACAACAACAGCAGCATGACCTATGAA 444

      304 LysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLy 320
          ||||||||||||:::::|||   |||||||||:::   |||:::::::::
      445 AAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATTCTGGAGCG 494

      320 sProAspIleSerGlnArgLeuValTyrGlnPhe 344
          :  |||  ::::::||||||||||||:::|||
      495 TGTGGAT...GGACGAAGACTGGTATATAAATTT 525

Seq_name: qb_est98:BG304069

seq_documentation_block:
LOCUS        BG304069      686 bp    mRNA            EST       23-FEB-2001
DEFINITION   tf66a04.x1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
             3815410 3' similar to TR:Q9YHW7 Q9YHW7 ETS DOMAIN PROTEIN ., mRNA
             sequence.
ACCESSION    BG304069
VERSION      BG304069.1  GI:13101596
KEYWORDS     EST.
SOURCE       zebrafish.
  ORGANISM   Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE    1  (bases 1 to 686)
  AUTHORS    Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
             Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
             Martin,J., Pape,D., Stoptoe,M., Underwood,K., Theising,B., Ritter
             ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE      Washu zebrafish EST Project 1999
  JOURNAL    Unpublished (1999)
COMMENT      Contact: S.L. Johnson
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
             Sequencing by: Washington University Genome Sequencing Center Clone
             distribution information can be found through the I.M.A.G.E.
             Consortium/LLNL, send email to: info@image.llnl.gov
             Seq primer: T7 from Gibco
             High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1..686
                     /organism="Danio rerio"
                     /strain="AB"
                     /db_xref="taxon:7955"
                     /clone="3815410"
                     /clone_lib="Sugano Kawakami zebrafish DRA"
                     /sex="mixed (one male and one female, including
                     unfertilized eggs)"
```

```
                     /dev_stage "adult"
                     /lab_host "DH10B (phage resistant)"
                     /note "Vector: pME18S-FL3; Site_1: BraII
                     (CACTGTGTG); Site_2: BraII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a BraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct BraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert. Size selection was performed to exclude fragments
                     <1.5kb. Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science) and
                     kindly donated by Dr. Koichi Kawakami. Custom primers for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA. "
BASE COUNT       161 a    182 c    174 g    129 t
ORIGIN


alignment_scores:
            Quality:  208.00              Length:      85
              Ratio:  3.302                 Gaps:       1
 Percent Similarity:  71.118   Percent Identity: 51.765

alignment_block:
 US-09-126-945b-2 x AI304069/rev   ..

 Align seq 1/1  to reverse of: AI304069  from: 1  to: 686

      251 LeuTrpGlnPheLeu.....LysGluLeuLeuLeuLysProHisSerTy 286
          ||||||||||||||||          ::::::|||              ::
      686 CTGTGGCAGTTTCTGGTGACGCATGGGATGATCTC............CA 646

      265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleG 282
          ::|||::|||||   |||   ::::::|||      :::   ||||::
      645 cGGCCATTTCATTGCCTGGACTGCACGTGAAATGGAGCTC...AAACTTA 599

      282 luAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgPro 298
            :::      :::|||||||||   ||||||||||:::|||||||||||
      598 TTGAGCCAGAGGAGGTGGCTCGACGCTGGGGCATCCAGAAGAACAGACCC 549

      299 AlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLy 315
          ||||||||||||||||||||||||||||::|||   ||||||||:::|
      548 GCCATGAACTATGACAAGCTGAGTCGCTCACTGCGCTACTACTATGAGAA 499

      315 sGlyIleIleArgLysProAspIleSer...GlnArgLeuValTyrGlnP 341
          |||||:|::::::|||     ::::::   ::::||   ||||||:::|
      498 GGGAATCATGCAGAAGGTAAAGGTTGCCGGCCAGAGGTACCTCTACAAAT 449

      341 heVal 332
          |||||
      448 TTGTG 444

Seq_name: gb_est26:AI891569

seq_documentation_block:
LOCUS        AI891569        483 bp    mRNA            EST       27-JUL-1999
DEFINITION   ng41q06.x1 Soares mouse kidney mkia Mus musculus cDNA clone
             IMAGE:2099962 3' similar to TR:P23274 P23274 BIG H-3 PRECURSOR
             ;, mRNA sequence.
ACCESSION    AI891569
VERSION      AI891569.1  GI:5597471
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 483)
  AUTHORS    Marra,M., Hillier,L., Kucaba,T., Martin,J., Rock,C., Wylie,T.,
             Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
             ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schark,K., Ritter
             ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
             Waterston,R. and Wilson,R.
  TITLE      The WashU-NCI Mouse EST Project 1999
  JOURNAL    Unpublished (1999)
COMMENT      other_FSTs: ng41q06.y1
             Contact: Marra M/WashU NCI Mouse EST Project 1999
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:996894
             Possible reversed clone; similarity on wrong strand
             Seq primer: custom primer used
             High quality sequence stop: 443.
FEATURES             Location/Qualifiers
     source          1..483
                     /organism "Mus musculus"
                     /strain "C57BL"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:2099962"
                     /clone_lib "Soares mouse kidney mkia"
                     /sex "female"
                     /dev_stage "adult"
                     /lab_host "DH10B"
                     /note "Organ: kidney; Vector: pME18S-FL3; Site_1: BraII
                     (CACTGTGTG); Site_2: BraII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a BraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct BraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert. Size selection was
                     performed to exclude fragments <1.5kb. Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science). Custom primers for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT        95 a    122 c    100 g    165 t      1 others
ORIGIN


alignment_scores:
            Quality:  207.50              Length:      82
              Ratio:  3.097                 Gaps:       1
 Percent Similarity:  81.707   Percent Identity: 45.122

alignment_block:
 US-09-126-945b-2 x AI891569/rev   ..

 Align seq 1/1  to reverse of: AI891569  from: 1  to: 483

      250 HisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGl 266
          ||||||||||:::|||:::::::::::::::|||||:::|||   :::
      447 CACTTATGGGAGTTCATCCGAGACATTCTCTTGAGCCCAGACAAGAACCC 398

      266 yArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluA 283
            :::|||:::|||   :::::::::::|||||||||||::::::
      397 AGGACTGATCAAATGGGAAGACCGTTCGGAAGGCATCTTCAGGTTCCTGA 348

      283 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299
          |||      |||||||:::|||||||||   :::|||||:::   :::
      347 AGTCAGAAGCTGTGGCTCAGCTGTGGGGGAAAAAGAAAAATAACAGTAGC 298

      300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
          |||:::|||:::|||||||||||:::::::|||   |||||||||:::
      297 ATGACATACGACAAGCTCAGCCGCCGCATATCAGATATACTACAAACGCA 248

      316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          |||:::::::::   |||   :::::|||||||||||||:::|||
      247 AATTCTGTAACGTTGTCGAT   CGACGACGGCTGGTCTATAACTTT 205

Seq_name: gb_est84:BF148041
```

```
seq_documentation_block:
LOCUS        BF148041       584 bp    mRNA            EST       26-OCT-2000
DEFINITION   nw93e06.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
             IMAGE:3469570 3' similar to TR:O70273 O70273 ETS HOMOLOGOUS FACTOR
             ;, mRNA sequence.
ACCESSION    BF148041
VERSION      BF148041.1  GI:11029436
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 584)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      other ESTs: nw93e06.y1
             Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:1088930
             High quality sequence stop: 488.
FEATURES             Location/Qualifiers
     source          1..584
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3469570"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       116 a    142 c    128 g    198 t
ORIGIN


alignment scores:
            Quality:  207.50            Length:     82
              Ratio:  3.097               Gaps:      1
 Percent Similarity:  81.707   Percent Identity: 45.122

alignment block:
 US-09-126-945B-2 x BF148041/rev   ..

 Align seq 1/1  to reverse of: BF148041  from: 1  to: 584

      250 HisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGl 266
          ||||||||||::|||::::::::::::::||||||::|||   :::
      448 CACTTATGGGAGTTCATCCGAGACATTCTCTTGAGCCCAGACAAGAACCC 399

      266 yArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluA 283
            :::|||::|||   :::::::::::::|||||||||:::::
      398 AGGACTGATCAAATGGGAAGACCGTTCGGAAGCCATCTTCAGGTTCCTGA 349

      283 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299
            |||       |||||:::|||||||||   :::||||||:::   :::
      348 AGTCAGAAGCTTGTGGCTCAGCTGTGGGGGAAAAAGAAAAATAACAGTAGC 299

      300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
          |||:::|||:::|||||||||||||:::::|||   ||||||||:::
      298 ATGACATACGACAAGCTCAGCCGGCTATGAGATATTACTACAAACGAGA 249

      316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
```

```
          |||:::::::::   |||   ::::::||||||||||||||:::|||
      248 AATCCTGGAACGTGTGGAT...GGACGACGGCTGGTCTATAAGTTT 206

seq_name: qb_est84:BF148053

seq_documentation_block:
LOCUS        BF148053       586 bp    mRNA            EST       26-OCT-2000
DEFINITION   nw93e07.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
             IMAGE:3469573 3' similar to TR:O70273 O70273 ETS HOMOLOGOUS FACTOR
             ;, mRNA sequence.
ACCESSION    BF148053
VERSION      BF148053.1  GI:11029448
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 586)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      other ESTs: nw93e07.y1
             Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:1088933
             High quality sequence stop: 489.
FEATURES             Location/Qualifiers
     source          1..586
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3469573"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I  oligo(dT) primer. Double stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       118 a    142 c    128 g    198 t
ORIGIN


alignment scores:
            Quality:  207.50            Length:     82
              Ratio:  3.097               Gaps:      1
 Percent Similarity:  81.707   Percent Identity: 45.122

alignment block:
 US-09-126-945B-2 x BF148053/rev   ..

 Align seq 1/1  to reverse of: BF148053  from: 1  to: 586

      250 HisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGl 266
          ||||||||||::|||::::::::::::::||||||::|||   :::
      448 CACTTATGGGAGTTCATCCGAGACATTCTCTTGAGCCCAGACAAGAACCC 399

      266 yArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluA 283
            :::|||::|||   :::::::::::::|||||||||:::::
      398 AGGACTGATCAAATGGGAAGACCGTTCGGAAGCCATCTTCAGGTTCCTGA 349

      283 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299
            |||       |||||:::|||||||||   :::||||||:::   :::
      348 AGTCAGAAGCTTGTGGCTCAGCTGTGGGGGAAAAAGAAAAATAACAGTAGC 299

      300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
```

```
          ||||:::|||:::|||||||||||||||:::::|||    |||||||||||:::
      298 ATGACATACGAGAAGCTCAGCCGGGCTATCAGATATTACTACAAACGAGA 249

      316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 441
            |||:::::::::    |||    ::::::|||||||||||||||||
      248 AATCCTGGAACGTGTGGAT...GGACGACGGCTGGTCTATAAGTTT 206

seq_name: qb_est84:BF117201

seq_documentation_block:
LOCUS       BF117201       600 bp    mRNA            EST       29-DEC-2000
DEFINITION  uy96e11.x1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:36674
            similar to TR:Q70273 Q70273 ETS HOMOLOGOUS FACTOR ;, mRNA sequence.
ACCESSION   BF117201
VERSION     BF117201.1  GI:10986677
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: uy96e11.y1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml

            MGI:1428204
            Seq primer: -40UP from Gibco
            High quality sequence stop: 426.
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Mus musculus"
                     /strain "C57/B6"
                     /db_xref "taxon:10090"
                     /clone "IMAGE:36674"
                     /clone_lib "NCI_CGAP_Mam5"
                     /tissue_type "tumor, gross tissue"
                     /dev_stage "7 months"
                     /lab_host "DH10B"
                     /note "Organ: mammary; Vector: pCMV SPORT6; Site_1: Sal;
                     Site_2: NotI; Cloned unidirectionally.  Primer: oligo dT.
                     Library constructed by Life Technologies.  Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT      135 a    151 c    115 g    199 t
ORIGIN


alignment_scores:
            Quality:  207.50             Length:     82
              Ratio:  3.097                Gaps:      1
 Percent Similarity:  81.707   Percent Identity: 45.122

alignment_block:
 US-09-126-945B-2 x BF117201/rev  ..

 Align seq 1/1  to reverse of: BF117201  from: 1  to: 600

      250 HisGlnTrpGlnPheLeuLysGluLeuGlnLeuLysProHisSerTyrG; 266
          ||||||||||:::|||||||||||||||||||:|||    :::
      543 CACCTATGGCAGTTCATCGCAGACATTCTCTTGAGCGCAGACAAGAAYYC 494

      266 yArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGlnA 284
```

```
          :::|||:::|||    ::::::::::::::||||||||:::::
      494 AGGACTGATCAAATGGGAAGACGGTTCGGAAGGCATCTTCAGTTCCTGA 444

      283 qPSerAlaGlnValArgLeuTrpGlyIleArgLysAsnArgProAla 299
            |||      |||||||:::|||||||||   :::|||||||:::   :::
      443 AGTCAGAAGCTGTGGCTCAGCTGTGGGGGAAAAAGAAAAATAACAGTAGC 394

      300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
          |||:::|||:::|||||||||||||||:::::|||   ||||||||||:::
      393 ATGACATACGAGAAGCTCAGCCGGGCTATCAGATATTACTACAAACGAGA 344

      316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
            |||:::::::::   |||   ::::::|||||||||||||||:::|||
      343 AATCCTGGAACGTGTGGAT...GGACGACGGCTGGTCTATAAGTTT 301

seq_name: qb_est100:BG388964

seq_documentation_block:
LOCUS       BG388964      1089 bp    mRNA            EST       12-MAR-2001
DEFINITION  602414721F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4523213 5',
            mRNA sequence.
ACCESSION   BG388964
VERSION     BG388964.1  GI:13282410
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1089)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10425  row: i  column: 06
            High quality sequence stop: 709.
FEATURES             Location/Qualifiers
     source          1..1089
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone "IMAGE:4523213"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host "DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI, Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      401 a    239 c    254 g    195 t
ORIGIN


alignment_scores:
            Quality:  207.50             Length:     88
              Ratio:  3.144                Gaps:      3
 Percent Similarity:  75.000   Percent Identity: 48.864

alignment_block:
 US-09-126-945B-2 x BG388964  .

 Align seq 1/1  to  BG388964  from  1  to: 1089

      246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuGluLeuLysPr 292
          ||:::  |||:::|||||:::|||||      |||||
       40 GCAAACACAATTTATCTTTGCCAGTTTTTACTGGCACTGCTCCAGGACAA 89
```

```
      262 oHisSerTyrGly.ArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
          :::          ::::::|||::|||   ::::::|||||||||||
       90 GGCTACTTGTCCTAAAGTACATCAACTGGACCCAGCCAGAGAAAGGCATT 139

      279 PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLy 295
          |||||||::   ||||||      |||::::|||||||||||   :::||
      140 TTTAAATTGGTGGATTCTAAAGCAGTGTCCAGGTTGTGGGGGAAGCACAA 189

      295 sAsn.ArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGln 311
          |||| :::|||   ||||||||||::   ::::::|||:::::|||
      190 AAACGAAACCTGATATCAATTATGAGACCATGGGAAGACCACTCAGGTAC 239

      312 TyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVa 328
          ||||||:::::|||||||::   |||   :::   :::||||||||||
      240 TATTACCAAAGGGGTATCTGGCAAAAGTGGAA...GGTCAGCGCTTGGT 286

      328 TTyrGlnPhe 336
          ||||||||||
      287 GTATCAGTT 296

seq_name: qb_est29:AL527893

seq_documentation_block:
LOCUS        AL527893          750 bp    mRNA             EST       13-FEB-2001
DEFINITION   AL527893 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC027YP16 5
             prime, mRNA sequence.
ACCESSION    AL527893
VERSION      AL527893.1  GI:12791386
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 750)
  AUTHORS    Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE      Full length cDNA libraries and normalization
  JOURNAL    Unpublished (2001)
COMMENT      Contact: Genoscope
             Genoscope  Centre National de Sequencage
             BP 191 91006 EVRY cedex  France
             Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1..750
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone "CS0DC027YP16"
                     /clone_lib "LTI_NFL003_NBC3"
                     /sex "male"
                     /tissue_type "neuroblastoma cells"
                     /lab_host "DH10B"
                     /note "Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI oligo(dT) primer. Five prime end
                     enriched, double stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 439
                     8471 Email : fliang@lifetech.com URL :
                     http://www.lifesearch.com"
BASE COUNT       226 a    173 c    178 g    167 t      6 others
ORIGIN


alignment scores:
            Quality:  206.50              Length:     248
              Ratio:  1.721                 Gaps:       7
 Percent Similarity:  48.387   Percent Identity: 22.984

alignment block:
 HS_09_126_945B_2 x AL527893    ..

Align seq 1/1  to: AL527893  from: 1  to: 750

       99 GlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGl 115
          ::::::|||::::::::::|||||||||::      ::::::      ::
      149 GAGGGACCCACGAACCAGCTGGATCTGCTCATCCGGGCCGTGGAAGCATC 198

      115 nGluHisSer.....................LeuGluGlnValG 123
          :   ||||||                     :::|||   ::::
      199 AGTTCACAGCAGTTATCCACACTGTACAGATAAGACAATTGAAGCTGCTG 248

      123 InSerMetValValGlyGlu.........ValLeuLysAsp..... 133
          ::::::::::      |||          ||||:|||
      249 AAGCCCTGCTTCATATGGAATCTCCTACCTGCTTGAGGGATTCAAAAGT 298

      134 .....................................IleGluThrAl 137
                                                   ||||:
      299 CCTGAATTCATCCATGCTGCTATGAGGCCAGATGTCATTACAAAAGCTGT 348

      137 acysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerA 154
          :   ::::::   :::      :::::|||||||||   ||||||
      349 AGTGGAGGTGTCAACTGAAGAGTCTGAACCCATGGATACCTCTCCT.... 394

      154 snValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProPro 170
                                                  ::::||
      395 .....................................ATTCCAACA 403

      171 MetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSe 187
                :::   :::   :::      ||:::         :::::
      404 TCACCAGATAGCCATGAACCAATGAAAAAAAAAAAA.....GTTGG 444

      187 rGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuHisA 204
          :::::::          ||:::||||||:::::::::
      445 CCGTAAACCAAAGACCCAGCAATCACCAATTTCCAATGGG..... 484

      204 laHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSer 220
                          :::           ::::||::::::
      485 ................TCTCCTGAGTTAGGTATAAAGAAGAAACCAAGA 517

      221 ProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerIrpThrAs 237
              |||
      518 GAAGGA..................................... 523

      237 pSerGlnValAspSerSerysSerGlyGlnProIleHisLeuTrpGlnP 254
                              :::|||:::      :::||||||:::|
      524 ....................AAAGGAAACACAACCTATTTGTGGGAGT 551

      254 heLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArg 270
          |||||   :::|||||||         ::::::      ||||:::|||::
      552 TTCTTTAGATCTACTTCAAGATAAAAATACCTTGTCCCAGGTATATTAAA 601

      271 IrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVa 287
          |||   :::::::|||||||||||||||::::   ||||||      ||
      602 TGGACTCAGAGACAAAAAGCCATATTCAAGCTGGTGGATTCAAAGGCTGT 651

      287 lAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspL 304
          |::::::|||||||||   :::||||||:::|||   ||||||||||:::
      652 CTCTAAGCTTTGGGCAAAGCATAAGAACAAACCAGACATGAACTATGAAA 701

      304 ysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIle 348
          ::::::|||::::::|||   ||||||:::::|||||||:::
      702 CCATGGGACCAGCTTGAGATACTACTACCAAAGGGAATTCTT 745

seq_name: qb_est79:BE778347

seq_documentation_block:
LOCUS        BE778347          665 bp    mRNA             EST       20-OCT-2000
DEFINITION   601463748F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3867130 5',
             mRNA sequence.
ACCESSION    BE778347
VERSION      BE778347.1  GI:10159545
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 665)
  AUTHORS   NIH MGC http://www.ncbi.nlm.nih.gov/
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9612  row: p  column: 11
            High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1..665
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3867130"
                     /clone_lib:"NIH_MGC_67"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: oligo dT.
                     Average insert size 1.75 kb.  Library constructed by Life
                     Technologies."
BASE COUNT      164 a    156 c    184 g    163 t
ORIGIN


alignment_scores:
            Quality:  206.00             Length:     78
              Ratio:  3.377                Gaps:      3
 Percent Similarity:  78.205   Percent Identity: 56.410

alignment_block:
 US-09-126-945B-2 x BE778347   ..

 Align seq 1/1  to: BE778347  from: 1  to: 665

     244 CysSerGlyGlnProIleHis LeuTrpGlnPheLeuLysGluLeuLeu 289
         |||   ||||||      ||| ||||||||||||||||   :::|||||
     266 TGCTATGGACAGTGCTATCACCCTGTGGCAGTTCCTTCTTCAGCTCCTGC 315

     290 euLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLys 276
         ||||||||||      ::::::|||   |||   :::::::::
     316 AGAAGCCTCAGAAC...AAGCACATGATCTGTTGGACCTCTAATGAT... 359

     277 GlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIl 293
         |||   ||||||:::   ::::::   :::|||||||||||||||||||
     360 GGGCAGTTTAAGCTTTGCAGGCAGAAGAGGTGGCTCGTCTCTGGGGGAT 409

     293 eArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleA 310
         ||||||||||||:::|||   |||||||||||||||||||||||||:::::|
     410 TCGCAAGAACAAGCCTAACATGAATTATGACAAACTCAGCCGAGCCCTCA 459

     310 rgGlnTyrTyrLysLysGly!IeIleArgLys 320
         ||   ||||||   |||:::|||||||:::||
     460 GATACTATTATGTAAAGAATATCATCAAAAAA 491

seq name: gb_est86:BF320950

seq_documentation_block:
LOCUS       BF320950        602 bp    mRNA             EST       29 FEB 2000
DEFINITION  uz57e08.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3673190 5'
            similar to SW:FRM_HUMAN P41161 ETS-RELATED PROTEIN FRM >, mRNA
            sequence.
ACCESSION   BF320950
VERSION     BF320950.1  GI:11269983
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov/image/html/ordering.shtml

            MGI:1433958
            High quality sequence stop: 437.
FEATURES             Location/Qualifiers
     source          1..602
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3673190"
                     /clone_lib="NCI_CGAP_Mam6"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      142 a    169 c    158 g    132 t      1 others
ORIGIN


alignment_scores:
            Quality:  205.50             Length:     85
              Ratio:  3.211                Gaps:      4
 Percent Similarity:  75.294   Percent Identity: 54.118

alignment_block:
 US-09-126-945B-2 x BF320950   ..

 Align seq 1/1  to: BF320950  from: 1  to: 602

     249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
         ::::::||||||||||||||      ||||||      |||   :::
     174 CTGCAGCTGTGGCAGTTCCTTCTTCAGCTCCTGGATGACCCA...GCCAA 220

     265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLys! 281
         ::::::||||||   |||   :::          ||:::   ||||||:
     221 TGCCCATTTCATCGCCTGGACTGCC      CCGGGCATGCAATTTAAAC 264

     281 leGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
         ::   :::      :::||||||||||   ||||||||||:::||||||||
     265 TCATAGAACCGCAAGAGGTTGCTCGCCGTTGGGGTATCCAGAAGAATCGG 314

     298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 314
         ||||||||||||||||||||||||||||||||||:::|||   ||||||::
     315 CCAGCCATGAACTATGACAAGCTGAGCGCTCTCTCCGCTATTACTATGA 364

     314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
         :|||||||||:::::|||          :::::|||   ||||||:::|
     365 AAAGGGCATCATCCAGAAGCCC   GCTGCGGAACGCTACGTCTACAAAT 411

     331 heVal 332
         |||||
     412 TTGTG 416
```

```
Seq_name: qb_est86:BE444411

seq_documentation_block:
LOCUS       BE444411       970 bp    mRNA            EST       22-NOV-2000
DEFINITION  602014802F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4150584
            5', mRNA sequence.
ACCESSION   BE444411
VERSION     BE444411.1  GI:11291697
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 970)
  AUTHORS   NIH MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9414  row: e  column: 01
            High quality sequence stop: 727.
FEATURES             Location/Qualifiers
     source          1..970
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4150584"
                     /clone_lib="NCI_CGAP_Brn64"
                     /tissue_type="glioblastoma with EGFR amplification"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: oligo dT.
                     Average insert size 1.52 kb. Constructed by Life
                     Technologies. Note: this is a NCI CGAP Library."
BASE COUNT      210 a    259 c    416 g    185 t
ORIGIN


alignment_scores:
            Quality:  205.50            Length:     84
              Ratio:  3.211               Gaps:      4
 Percent Similarity:  77.108   Percent Identity: 54.217

alignment_block:
 US-09-126-945B-2 x BE444411   ..

 Align_seq 1/1  to: BE444411  from: 1  to: 970

      249 IleHisLeuTrpGluPheLeuLysGluLeuLeuLysProHisSer-Tyr 265
          ::::::!!!!!!!!!!!!!!!!      !!!!!!      !!!   :::
      168 CTGCAGGTGTGGCAATTTCTGGTGGCCTTGCTGGATGACCCA...ACAAA 214

      265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLys1 281
          ::::::!!!!!!   !!!   :::     :::!!!::   !!!!!!:
      215 TGCCCATTCATTGCCTGGACGGGC.....CGGGGAATGCAGTTCAAGC 258

      281 IleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
          ::   :::      :::!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
      259 TCATTGAGCCTGAGGAGGTGGCCAGGCTCTGGGGCATCCAGAAGAACCGG 308

      298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 314
          !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!:::!!!   !!!!!!:
      309 CCAGCCATGAATTACGACAAGCTGAGCCGCTCGCTCCGATACTATTATGA 358

      314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGln 333
          :!!!!!!!!!:::::!!!         :::::!!!   !!!!!!:::
      359 GAAAGGCATCATGCAGAAGGTG...GCTGGTGAGCGTTACGTGTACAAG 404

Seq_name: qb_est69:BE026246

seq_documentation_block:
LOCUS       BE026246       479 bp    mRNA            EST       29-JUN-2000
DEFINITION  db34c04.y1 Blackshear/Soares normalized Xenopus eqq library Xenopus
            laevis cDNA clone IMAGE:3400390 5' similar to qb:L19953 Mus
            musculus EKP mRNA, complete cds (MOUSE)., mRNA sequence.
ACCESSION   BE026246
VERSION     BE026246.1  GI:8319607
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Bento Soares and M. Fatima Bonaldo
            (University of Iowa). DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer:  40RP from Gibco
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1..479
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:3400390"
                     /clone_lib="blackshear/Soares normalized Xenopus eqq
                     library"
                     /sex="female"
                     /tissue_type="unfertilized eqq"
                     /cell_type="unfertilized eqq"
                     /dev_stage="unfertilized eqq"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3 Pac; Site_1: EcoRI; Site_2: NotI;
                     PolyA selected mRNA was prepared from unfertilized Xenopus
                     laevis eqqs.  The library was constructed in the vector
                     pT7T3 Pac as described in Bonaldo, M.F., Lennon, G. and
                     Soares, M.B. 'Normalization and subtraction:  two
                     approaches to facilitate gene discovery',  Genome Research
                     6:791 806, 1996.  The first strand synthesis used a
                     NotI-dT18 primer; double stranded cDNAs were ligated to
                     EcoRI adapters, digested with NotI, and directionally
                     cloned into the NotI and EcoRI digested pT7T3-Pac vector.
                     The library contained approximately 7.2 X 10^5
                     recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      141 a    108 c    131 g     99 t
ORIGIN


alignment_scores:
            Quality:  205.00            Length:    163
              Ratio:  2.158               Gaps:      7
 Percent Similarity:  58.282   Percent Identity: 44.742

alignment_block:
 US-09-126-945B-2 x BE026246   ..
```

```
Align seq 1/1  to: BE026246  from: 1  to: 479

     170 ProMetGlyLysAlaPheGlnGluLeuAlaGly...LysGluLeuCysAl 185
         |||   |||||||:   |||:::   ::::::   |||      ||||:
      84 CCTAGCGGTCGATCCCGTCAGCAAGAAGGAAACATGAAAACTGGCTGTAG 83

     185 aMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspVal 202
         :::::::|||:::                 |||   ||||||
      84 CATTGGGCAAAAG                 GCCAAA GCAGGC        1 8

     202 euHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArg 219

     108 ..................................................  108

     219 ThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerIr 245
               ||:::   |||:::                       :::||
     109 .........GGAGGTTATCACTTC.................CCAGACTG 141

     235 pThrAspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuI 252
         |:::      :::::::|||||   :::::::::   ||||:::|||
     132 GGTG...TACAAGGCGGAGTCAAGTCCCGGCTCCAGACAGATCCAGCTGT 178

     252 rpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPhe 268
         |||:::|||::   |||||||||   |||   :::::::::   :::
     179 GGCATTTTATTTTGCAACTGCTGCAGAAC . GACGAGTTCCACATGTT 225

     269 IleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAl 285
         |||   |||   :::   |||   |||   |||   ||||:|||
     226 ATTGCCTGGCAGCACGGGCAATATCCACAGTTTGTCATTAAACACCCACA 275

     285 aGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnI 302
          :::|||||||||||||||||||   ||||||   :::|||   |||||||
     276 TGAGGTGGCCAGGCTATGGGGCCGGCGCAAATGCAAACCTCAGATGAACT 325

     302 yrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIle 319
         |||||||||||||||||:::::|||   ||||||:::|||   ||||::
     326 ACGACAAGCTGAGCCGAGCTCTGAGATACTATTACAACAAGGCAGATCCTG 375

     319 ArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
         :::|||        :::::::|||:::::::|||:::|||
     376 CATAAAACAAAA...GGAAAGCGTTTTACTTACAAATTC 411

seq_name: qb_est87:BF446190

seq_documentation_block:
LOCUS       BF446190      625 bp    mRNA            EST       01-DEC-2000
DEFINITION  7p33f08.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3647751 3'
            similar to TR:O95238 O95238 EIS TRANSCRIPTION FACTOR PREF. ;, mRNA
            sequence.
ACCESSION   BF446190
VERSION     BF446190.1  GI:11511328
KEYWORDS    EST
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael K.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1..625
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /clone="IMAGE:3647751"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host "DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clonelDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      113 a    174 c    202 g    135 t      1 others
ORIGIN


alignment_scores:
            Quality:  205.00              Length:     47
              Ratio:  4.659                 Gaps:      0
 Percent Similarity:  93.617   Percent Identity: 93.617

alignment_block:
 HG-09-126-945B-2 x BF446190/rev  ..

 Align seq 1/1  to reverse of: BF446190  from: 1  to: 625

     289 ArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLe 305
         ||||||||||||||||||||   ||||||||||||||      |||||
     625 CGGCTGTGGGGCATCCGCAACANCCGTCCCGCCATGAACTA.GCCAAGCT 577

     305 uSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProA 322
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     576 GAGCCGCTCCATCCGCCAGTATTACAACAAGGGCATCATCCGGAAGCCAG 527

     322 spIleSerGlnArgLeuValTyrGlnPheValHisProIle 335
         |||||||||||||||||||||||||||||||||||||||
     526 ACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACCCCATT 486

seq_name: qb_est87:BF440177

seq_documentation_block:
LOCUS       BF440177      531 bp    mRNA            EST       01-DEC-2000
DEFINITION  BS29000192203 Lewin Cattle Spleen Bos taurus cDNA clone
            BS29000192203 3', mRNA sequence.
ACCESSION   BF440177
VERSION     BF440177.1  GI:11500060
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Band,M.R., Larson,J.H., Rebeiz,M., Green,C.A., Heyen,D.W., Donovan
            ,J., Windish,R., Steining,C., Mahyuddin,P., Womack,J. and Lewin
            ,H.A.
  TITLE     An ordered comparative map of the cattle and human genomes
  JOURNAL   Genome Res. 10 (9), 1359-1368 (2000)
  MEDLINE   20442530
COMMENT     Contact: Lewin, H. A.
            W. M. Keck Center for Comparative and Functional Genomics
            University of Illinois at Urbana-Champaign
            340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
            61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5617
```

```
            Email: h lewin@uiuc.edu
            Funding for Cattle EST sequencing was provided by the USDA National
            Research Initiative, Project No. 98-35205-6644, and a grant from
            the Japanese Ministry of Agriculture Fisheries and Forestry to
            H.A.Lewin and J.E.Womack. Base calling/Quality scores: PHRED orm
            Washington University Genome Center.Vector trimming: Cross_Match
            from Washington University Genome Center PHRAP suite. This sequence
            is vector free and at least 150bp in length.
            Insert Length: 531   Std Error: 0.00
            Seq primer: TACGACTCACTATAGGGCGAAT
            High quality sequence stop: 531.
FEATURES             Location/Qualifiers
     source          1..531
                     /organism "Bos taurus"
                     /strain "Angus"
                     /db_xref "taxon:9913"
                     /clone "BS2900192203"
                     /clone_lib "Lowin Cattle Spleen"
                     /sex "Female"
                     /dev_stage "Adult"
                     /note "Organ: Spleen, Vector: pBluescript SK(+/-)"
BASE COUNT        92 a    153 c    145 g    137 t      4 others
ORIGIN


Alignment scores:
            Quality:  204.50            Length:    191
              Ratio:  2.153               Gaps:     10
 Percent Similarity:  49.738   Percent Identity: 42.461

Alignment block:
 US-09-126-945B-2 x BF440177/rev   ..

 Align seq 1/1  to reverse of: BF440177  from: 1  to: 531

     149 LysLeuLeuAsnIleThrAlaAspProMetAsp...TrpSerProSerAs 154
         |||:::|||||:::      :::   :::|||   |||   |||
     528 AAAATCCTAAATCTCCAACAGGAGTCNCTGGACTCCTGGGTCCCCCAGGCC 479

     154 nValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProM 171
          :::   :::   |||   |||
     478 AATTCTTAGGTCCCTTAGATGG..................................... 457

     171 etGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSer 187
                                 ::::::::::|||   :::   :::
     456 ........................AGGAATCAGGATTGCTGGTCTCCGAG 430

     188 GluGluGlnPheArg..............GlnArgSerProLeuGlyGl 199
         ||||||||||:::|||              ::|||      ||||||
     429 GAGGAGCAGATCCGGATTCCTGGGTCATCGAAGCGACATCATCTGGGGCC 380

     199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetL 216
             |||   :::||||||
     379 AGGAAACCTGAGTGTACACTTG..................................... 358

     216 ysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGlu 232
                 ||||||        ||||||
     357 ..........TCCCCA.........CACTAC.................. 346

     233 GluSerTrpThrAspSerGluValAspSerSerCysSerGlyGln.Prot 249
                                                 ::: ||||
     345 ..........................................AGGPCCCA 338

     249 IeHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyr 265
         ||:::|||||||||||||||   |||||||||          |||:::
     337 TTCAGCTGTGGCAGTTCCTCCTGGAGCTGCTC.........CACGACGGA 297

     266 GlyArg......PheIleArgTrpLeuAsnLysGluLysGlyIlePheLy 280
         :::|||         |||||||||   ::::::::::::      |||::
     296 GCGCGTAGAAGCTGCATCCGCTGGACCGGCAATAGCCGAGAG...TTCCA 250

     280 sIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnA 297
         ::::   |||      :::||||||||||||||||||   ||||||::::

     249 ACTGTGCGACCCGAAAGAGGTGGCCGGCTGTGGGGTGAGCGCAAGAGCA 200

     297 rgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyr 313
         ::|||:::|||||||||||:::|||||||||||||:::   |||   ||||||
     199 AGCCAGGCATGAATTACGAGAAGCTGAGCCGAGGTCNCGGCTACTACTAC 150

     314 LysLysGlyIleIleArgLys 320
         ::::::   |||:::||||||
     149 CGCCGCGACATCGTGCGCAAG 129
```

```
OM of: US-09-126-945B-2 to  N_Geneseq_0601 *   <or_format   g1>

Date: Sep 18, 2001  7:12 AM

About: Results were produced by the GeneForm software  version 4.5
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters
  -MODEL=framer_p2s model -DEV=xlp
  -QS="sgn2 1 /usr1/sptmp/US09126945_run1_17[illegible]
  -DB N_Geneseq_0601 -pMI fastap -SUFFIX=mi -AF=32 -o=
  -GAPEXT=4.000  MINMATCH=0.100  [illegible]  -L=GAPEXT
  -QGAPOP=4.500 -QGAPEXT=6.000  XGAP=15.500  MGAPEXT=1.50
  -PGAPOP=6.000 -PGAPEXT=7.000 -VGAPOP=10.000 -YGAPEXT=0.500
  -DELOP 6.000 -DELEXT=7.000 -START 1 -MATRIX blosum62
  -IPANS human40.cdl -LIST=150 -DISPLAY=200 -THR_SCORE=[illegible]
  -THR_MAX 1.00 -THR_MIN    ALIGN=200  MODE 1=AL  OUTPUT=
  -NORM=0.0  MINLEN     MAXLEN 2000000000
  -USER US09126945  "CONT 1 20"  MPI=[illegible]  M  VIEW
   WAIT -HPFALG 1

Search Information block
Query: US-09-126-945B-2
Query length: 835
Database: N_Geneseq_0601 *
Database sequences: 240101
Database length: 61866809
Search time (sec): 111.98  0

score_list
Sequence          Std Score   Score   Expected  Length  Identity  Fitness
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA50651 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAY96561 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAY96560 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAE21638 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA01651 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA.94554 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA.94553 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA81654 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAC19651 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA81652 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA81653 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA94156 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA94157 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA.98656 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA.98657 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA.98658 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA.98659 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA.98660 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAA47450 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AA.18658 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA81655 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAY96562 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV55657 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA81656 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA81654 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AA.18657 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA81655 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAA96562 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA.94561 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA72415 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV20455 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV10701 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV06655 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA.94702 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV15620 *  1765 *  1765.5  0.0e+00  1064
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AA.18658 *  1765 *  1765.5  0.0e+00  1064

/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA85595 *  239.00  325.65  4.5e-16  848
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV08854 *  238.50  322.64  9.3e-16  1116
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA[illegible]  238.50  322.64  9.3e-16  1116
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAP15985 *  238.50  317.64  1.0e-09  1884
/SIDS8/geodata/geneseq/geneseqpo/NA1995.DAT:AAQ91769 *  236.00  314.59  1.5e-09  1800
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV91167  234.50  314.63  2.8e-10  852
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV58521 *  234.50  319.63  7.8e-10  852
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA96284  234.50  319.63  7.8e-10  852
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAA01646 *  232.50  311.42  2.2e-09  1716
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAA01645 *  232.50  307.09  4.9e-09  2744
/SIDS8/geodata/geneseq/geneseqpo/NA1996.DAT:AAY37987 *  230.50  305.62  4.7e-09  2664
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AA[illegible]  230.50  305.62  4.7e-09  2664
/SIDS8/geodata/geneseq/geneseqpo/NA1994.DAT:AAQ55149 *  230.50  305.58  4.7e-09  2678
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV42688 *  229.50  302.75  6.8e-09  2416
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAE50015 *  229.50  302.75  6.8e-09  2416
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAP54030 *  229.50  302.75  6.8e-09  2416
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAB94365 *  225.50  309.86  2.7e-09  1587
/SIDS8/geodata/geneseq/geneseqpo/NA1993.DAT:AAQ50662 *  225.00  298.79  1.1e-08  1980
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAT41849 *  225.00  297.06  1.4e-08  2200
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAY06608 *  225.00  297.06  1.4e-08  2200
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAE33871 *  206.50  277.47  2.3e-07  2376
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAB00769 *  204.00  262.54  4.4e-06  3608
/SIDS8/geodata/geneseq/geneseqpo/NA1997.DAT:AAI47198 *  194.50  252.99  4.0e-06  2997
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAP12704 *  192.50  252.19  4.5e-06  3[illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV10765 *  190.50  227.70  0.0002  3752
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV20472 *  170.00  225.23  0.0001  1934
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA81647 *  164.50  216.49  0.0004  1527
/SIDS8/geodata/geneseq/geneseqpo/NA1997.DAT:AAT47200 *  160.00  205.34  0.0038  2544
/SIDS8/geodata/geneseq/geneseqpo/NA1991.DAT:AAQ14187 *  155.00  204.68  0.0020  1313
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAP42329 *  154.00  209.94  0.0010  654
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAZ[illegible]  145.50  204.96  0.0009  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA44389 *  145.00  201.06  0.0021  449
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAF16096 *  109.00  139.50  8.45  1460
/SIDS8/geodata/geneseq/geneseqpo/NA1994.DAT:AAZ24399 *  109.00  139.50  8.46  1467
/SIDS8/geodata/geneseq/geneseqpo/NA1997.DAT:AAT96643 *  108.50  137.21  11.44  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA94836 *  108.50  137.21  11.44  1766
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAP21305 *  107.00  122.47  25.14  6552
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA35193 *  107.00  122.47  25.14  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAO25975 *  105.00  127.30  11.20  1329
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAO76988 *  [illegible]  [illegible]  [illegible]  1668
/SIDS8/geodata/geneseq/geneseqpo/NA1994.DAT:AAV12201 *  104.50  128.44  35.57  2438
/SIDS8/geodata/geneseq/geneseqpo/NA1997.DAT:AAA14624 *  104.50  128.44  35.57  2438
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA79042 *  104.50  126.74  43.40  2994
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AA[illegible]  [illegible]  124.68  [illegible]  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA1993.DAT:AAQ49024 *  103.00  134.82  15.41  997
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAZ23501 *  102.00  [illegible]  [illegible]  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAZ84857 *  102.50  122.94  70.72  4220
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAE62743 *  102.50  122.94  70.72  4220
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAZ79994 *  102.00  142.68  [illegible]  1[illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV55656 *  102.00  142.68  [illegible]  1089
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAZ[illegible]  101.50  121.48  [illegible]  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AA.42638 *  101.50  121.48  85.66  3244
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAZ[illegible]  101.50  121.48  [illegible]  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAV42657 *  101.50  121.48  85.66  3244
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAZ14156  101.50  121.48  [illegible]  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAZ20482 *  100.00  119.72  27.42  1142
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV10666 *  101.00  111.51  [illegible]  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA1996.DAT:AAN68104 *  100.50  121.44  [illegible]  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA2001.DAT:AAE[illegible]  100.50  120.21  [illegible]  1248
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA27141 *  100.50  124.73  56.19  1992
/SIDS8/geodata/geneseq/geneseqpo/NA1991.DAT:AAE14425 *  100.50  118.24  1[illegible]  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAV[illegible]  [illegible]  [illegible]  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAF21137 *  99.50  111.43  309.55  6944
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA35015 *  99.50  111.43  309.55  6944
/SIDS8/geodata/geneseq/geneseqpo/NA1992.DAT:AAQ28399 *  99.00  114.14  218.44  4854
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA[illegible]  98.50  [illegible]  [illegible]  [illegible]
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAA27147  98.50  120.53  96.26  2309
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAX27922 *  98.50  118.99  117.[illegible]  2714
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAV74156 *  98.00  110.49  349.10  6154
/SIDS8/geodata/geneseq/geneseqpo/NA1999.DAT:AAV08901 *  98.00  110.49  349.10  6154
/SIDS8/geodata/geneseq/geneseqpo/NA2000.DAT:AAB54177 *  98.00  110.49  349.10  6154
/SIDS8/geodata/geneseq/geneseqpo/NA1998.DAT:AAN60961 *  98.00  107.07  541.65  8880
```

/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAA62498	97.00	127.14	134.98	1252
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAC19665	97.00	127.14	134.98	5243
/SIDSR/geqdata/geneseq/geneseqp/NA1998:IAI:AAV24264	97.00	127.14	134.98	5243
/SIDSR/geqdata/geneseq/geneseqp/NA1999:IAI:AAV94308	97.00	127.14	134.98	6895
/SIDSR/geqdata/geneseq/geneseqp/NA2001:IAI:AAC86033	97.00	127.14	134.98	6895
/SIDSR/geqdata/geneseq/geneseqp/NA1997:IAI:AAI66961	96.50	125.22	130.17	1400
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAC70044	96.50	125.22	130.17	1400
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAC96781	96.50	125.22	130.17	1400
/SIDSR/geqdata/geneseq/geneseqp/NA1998:IAI:AAX25270	96.50	111.63	205.32	4202
/SIDSR/geqdata/geneseq/geneseqp/NA1999:IAI:AAV69717	96.50	111.63	272.11	4281
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAV67200	96.50	111.63	208.32	4281
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAC91170	96.00	111.06	204.28	1295
/SIDSR/geqdata/geneseq/geneseqp/NA1999:IAI:AAX87620	96.00	110.51	448.29	1992
/SIDSR/geqdata/geneseq/geneseqp/NA2001:IAI:AAC89557	96.00	105.46	673.79	9460
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAX98686	96.00	110.96	178.76	2567
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAA64341	95.50	110.07	199.17	2871
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAC68405	95.00	104.89	198.40	2894
/SIDSR/geqdata/geneseq/geneseqp/NA1992:IAI:AAZ25697	95.00	104.89	250.26	3490
/SIDSR/geqdata/geneseq/geneseqp/NA2001:IAI:AAS00022	95.00	104.89	473.40	6892
/SIDSR/geqdata/geneseq/geneseqp/NA1999:IAI:AAX87626	95.50	110.52	161.14	2299
/SIDSR/geqdata/geneseq/geneseqp/NA1998:IAI:AAV55365	95.00	104.89	69.60	1062
/SIDSR/geqdata/geneseq/geneseqp/NA2001:IAI:AAI71523	95.00	104.89	134.22	1917
/SIDSR/geqdata/geneseq/geneseqp/NA2001:IAI:AAC92270	95.00	104.89	134.22	1917
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAA49171	94.50	104.34	177.68	2298
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAA72738	94.50	104.34	236.44	3057
/SIDSR/geqdata/geneseq/geneseqp/NA1998:IAI:AAV53654	94.50	104.34	167.16	2161
/SIDSR/geqdata/geneseq/geneseqp/NA1998:IAI:AAV21867	94.50	104.34	76.62	991
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAL20478	94.00	103.77	76.62	991
/SIDSR/geqdata/geneseq/geneseqp/NA1998:IAI:AAV55368	94.00	103.77	80.67	1077
/SIDSR/geqdata/geneseq/geneseqp/NA1998:IAI:AAV11215	94.00	103.77	169.10	2238
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAA74191	94.00	103.77	209.87	2777
/SIDSR/geqdata/geneseq/geneseqp/NA2001:IAI:AAL01479	94.00	103.77	209.87	2777
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAC58089	94.00	103.77	643.51	5567
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAC58012	94.00	103.77	643.51	5567
/SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAC58010	94.00	103.77	643.51	5567
/SIDSR/geqdata/geneseq/geneseqp/NA1998:IAI:AAV57602	94.00	103.77	667.73	6004

Seq name: /SIDSR/geqdata/geneseq/geneseqp/NA2000:IAI:AAZ50691

seq documentation block:

```
ID   AAZ50691 standard; cDNA; 1894 BP.
XX
AC   AAZ50691;
XX
DT   23 MAY 2000  (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis; treatment; prostate cancer; blood coagulation;
KW   autoimmune disorder; haematopoietic; immuno/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             416..1423
FT                   /*tag= a
FT                   /product= "Human prostate derived Ets factor"
FT                   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999;  99WO-US17470.
XX
```

```
PR   31-JUL-1998;  98US-126945.
XX
PA   (GEMA-) HUMAN GENOME SCI INC.
PA   (BILH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Liberman HA, Seetharam RS, Fenoglio CA, Endress GA, Rosen CA;
XX
DR   WPI; 2000-227617/19.
DR   P-PSDB; AAY45004.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative -
XX
PS   Claim 4; Fig 1; 31pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (A.T.C.C. 75487). PDEF gene expression is highest in tissues involved in
CC   hormone sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,
CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   and antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in cancer therapies as factors, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders (like neoplasm and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration). They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 368 A; 654 C; 571 G; 402 T; 0 other;
```

alignment scores:

```
                Quality: 1785.00              Length:     855
                  Ratio: 5.328                  Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment block:

```
 US-09-126-945B-2 x AAZ50691    ..

 Align seq 1/1  to: AAZ50691  from: 1  to: 1894

       1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     416 ATGGGCAGCGCCAGCCCGGGTCTGAGCAGCGTATCCCCAGCCACCTCC 465

      17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     466 GCTGCCCCCGACACGGTGTCGCGGACAGGCTTGGAGAAGGCGGCAGCGG 515

      34 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     516 GGGCAGTGGGTCTCGAGAGACGGGACTGGAGTCCCAGTCCACCGCCACG 565

      51 ProGlnGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     566 CCCGAGCAGGGCCTGTCCGCCTTCTACCTCTCCTACTTTGACATGCTGTA 615

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     616 CCCTGAGGACAGCAGCTGGGCAGCCAAGGCCCCTGGGGCCAGCAGTCGGG 665

      84 luGluProProGluGluProGluGlnCysProValIleAspSerGluAla 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     666 AGGAGCCACCTGAGGAGCCTGAGCAGTGCCCGGTCATTGACAGCGCAAGCC 715
```

```
  101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHi 117
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  715 CCAGCGGGCAGCCTGCACTTGGTGCCCGGCGGGCTGACCTTGGAGGAGCA 765

  117 sSerLeuGlnGlnValGlnSerMetValValGlyGluValLeuLysAspI 134
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  766 CTCCCTGGACGAGGTGCAGTCCATGGTGGTGGGCGAAGTGCTCAAGGACA 815

  134 leGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrp 150
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  816 TCGAGACGGCCTGCAAGCTGCTCAACATCACCGCAGATCCCATGGACTGG 865

  151 SerProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrAr 167
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  866 AGCCCCAGCAATGTGCAGAAGTGGCTCCTGTGGACAGAGCACCAATACCG 915

  167 qLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuC 184
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  916 GCTGCCCCCATGGGCAAGGCCTTCCAGGAGCTGGCGGGCAAGGAGCTGT 965

  184 ysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAsp 200
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  966 GCGCCATGTCGGAGGAGCAGTTCCGCCAGCGCTCGCCCCTGGGTGGGGAT 1015

  201 ValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGl 217
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1016 GTGCTGCACGCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAAAGA 1065

  217 nArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluS 234
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1066 GCGGACTTCACCTGGGGCGATTCACTACTGTGCCTCGACCAGTCAGGAGA 1115

  234 erTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHis 250
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1116 GCTGGACCGACAGCGAGGTGGACTCATCATGCTCCGGGCAGCCCATCCAC 1165

  251 LeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyAr 267
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1166 CTGTGGCAGTTCCTCAAGGAGTTGCTACTCAAGCCCCACAGCTATGGCCG 1215

  267 qPheIleArgIrpLeuAsnLysGluLysGlyIlePheLysIleGluAspS 284
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1216 CTTCATTAGGTGGCTCAACAAGGAGAAGGGCATCTTCAAAATTGAGGACT 1265

  284 erAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMet 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1266 CAGCCCACGTGGCCCGGCTGTGGGGCATCCGCAAGAACCGTCCCGCCATG 1315

  301 AsnTyrAspLysLeuSerArgSerIleArgIleTyrTyrLysLysGlyIl 317
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1316 AACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAAGGGCAT 1365

  317 eIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheValHisP 334
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1366 CATCCGCAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACC 1415

  334 roIle 335
      ||||
 1416 CCATC 1420
```

Seq: Unnamed: US-09-126-945B-2.rng, 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;

```
Seq_documentation_block:
ID   AAC83261 standard; DNA; 1894 BP.
XX
AC   AAC83261;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor gene product in a biological
PI   sample from a cell -
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the DNA encoding GSEF.
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;
```

```
alignment scores:
            Quality: 1785.00              Length:     335
              Ratio:  5.328                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment block:
 US 09-126-945B-2 x AAC83261   ..

 Alignment 1/2   1   AAC83261   Query 1   to 1894

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      416 ATGGGCAGCGCCAGCCCGGGTCTGAGCAGCGTATCCCCCAGCCACCTCCT 465

       17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      466 GCTGCCCCCGACACGGTGTCGCGGACAGGCTTGGAGAAGGCGGCAGCGG 515

       34 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 516 GGGCAGTGGGTCTCGAGAGACGGGACTGGAGTCCCAGTCCACCCGCCACC 565

  51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 566 CCCGAGCAGGGCCTGTCCGCCTTCTACCTCTCCTACTTTGACATGCTGTA 615

  67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 616 CCCTGAGGACAGCAGCTGGGCAGCCAAGGCCCCTGGGGCCAGCAGTCGGG 665

  84 lnGluProProGlnGluProGluGlnCysProValIleAspSerGlnAla 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 666 AGGAGCCACCTGAGGAGCCTGAGCAGTGCCCGGTCATTGACAGCCAAGCC 715

 101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHi 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 716 CCAGCGGGCAGCCTGGACTTGGTGCCCGGCGGGCTGACCTTGGAGGAGCA 765

 117 sSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspI 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 766 CTCGCTGGAGCAGGTGCAGTCCATGGTGGTGGGCGAAGTGCTCAAGGACA 815

 134 leGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrp 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 816 TCGAGACGGCCTGCAAGCTGCTCAACATCACCGCAGATCCCATGGACTGG 865

 151 SerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrAr 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 866 AGCCCCAGCAATGTGCAGAAGTGGCTCCTGTGGACAGAGCACCAATACCG 915

 167 gLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuC 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 916 GCTGCCCCCCATGGGCAAGGCCTTCCAGGAGCTGGCGGGCAAGGAGCTGT 965

 184 ysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAsp 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 966 GCGCCATGTCGGAGGAGCAGTTCCGCCAGCGCTCGCCCCTGGGTGGGGAT 1015

 201 ValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGl 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1016 GTGCTGCACGCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAAAGA 1065

 217 uArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluS 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1066 GCGGACTTCACCTGGGGCGATTCACTACTGTGCCTCGACCAGTGAGGAGA 1115

 234 erTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHis 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1116 GCTGGACCGACAGCGAGGTGGACTCATCATGCTCCGGGCAGCCCATCCAC 1165

 251 LeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyAr 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1166 CTGTGGCAGTTCCTCAAGGAGTTGCTACTCAAGCCCCACAGCTATGGCCG 1215

 267 gPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspS 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1216 CTTCATTAGGTGGCTCAACAAGGAGAAGGGCATCTTCAAAATTGAGGACT 1265

 284 erAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMet 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1266 CAGCCCAGGTGGCCCGGCTGTGGGGCATCCGCAAGAACCGTCCCGCCATG 1315

 301 AsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIl 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1316 AACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAAGGGCAT 1365

 317 eIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheValHisP 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1366 CATCCGGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACC 1415

 334 roIle 335
     |||||
1416 CCATC 1420
```

Seq_name: [illegible]NA2001.DAT:AAC83266

Seq_documentation_block:

```
ID   AAC83266 standard; DNA; 3317 BP.
XX
AC   AAC83266;
XX
DT   16-MAR-2001  (first entry)
XX
DE   Gland specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW   Transcription factor; gland specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ss.
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999;   99US-0134112.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Kaufmann J,  Xin H,  Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49633.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland specific Ets transcription factor gene product in a biological
PT   sample from a cell
XX
PS   Disclosure; Fig 2; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as
CC   grading and staging of cancers by detecting GSEF expression in a
CC   biological test sample. The method may also be used to monitor patients
CC   having a predisposition to develop a particular cancer. GSEF polypeptides
CC   are useful for producing antibodies, in cancer diagnosis, prognosis,
CC   grading, staging and management of breast and prostate tumours, and in
CC   detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC   useful in gene therapy. GSEF gene product expression levels can be used
CC   in conjunction with any tissue in which an alteration in GSEF gene
CC   product expression levels is associated with development of a
CC   cancer associated phenotype. Cancers, which can be monitored include
CC   cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC   adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC   lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC   human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC   represents the cDNA encoding GSEF.
XX
SQ   Sequence 3317 BP; 710 A; 1026 C; 920 G; 611 T; 0 other;
```

alignment_scores:

```
            Quality:  1785.00            Length:    335
              Ratio:     5.328             Gaps:      0
 Percent Similarity:   100.000  Percent Identity:  100.000
```

alignment_block:

US-09-126-945B-2 x AAC83266 ..

Align_seq_1/1 to: AAC83266 from: 1 to: 3317

```
       1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe  17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1815 ATGGGCAGCGGCCAGCCCGGGTCTGAGCAGCGTATCCCCAGCCACCTCCT  1864

      17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG  34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1865 GCTGCCCCCGACACGGTGTCGCGGACAGGCTTGGAGAAGGCGGCAGCGG  1914

      34 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr  50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1915 GGGCAGTGGGTCTCGAGACGGGACTGGAGTCCAGTCCACCCGCCACG  1964

      51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy  67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1965 CCCGAGCAGGGCCTGTCGGCCTTCTACCTCTCCTACTTTGACATGCTGTA  2014

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG  84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2015 CCCTGAGGACAGCAGCTGGGCAGCCAAGCCCCCTGGGGCCAGCAGTCGGG  2064

      84 luGluProProGluGluProGluGlnCysProValIleAspSerGluAla  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2065 AGGAGCCACCTGAGGAGCCTGAGCAGTGCCCGGTCATTGACAGCGAAGCC  2114

     101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHi  117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2115 CCAGCGGGCAGCCTGGACTTGCTGCCCGGCGGCCTGACCTTGGAGGACCA  2164

     117 sSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspI  134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2165 CTCGCTGGAGCAGGTGCAGTCCATGGTGGTGGGCGAAGTGCTCAAGGACA  2214

     134 leGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrp  150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2215 TCGAGACGGCCTGCAAGCTGCTCAACATCACCGCAGATCCCATGGACTGG  2264

     151 SerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrAr  167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2265 AGCCCCAGCAATGTGCAGAAGTGGCTCCTGTGGACAGAGCACCAATACCG  2314

     167 gLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuC  184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2315 GCTGCCCCCCATGGGCAAGGCCTTCCAGGAGCTGGCGGGCAAGGAGCTGT  2364

     184 ysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAsp  200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2365 GCGCATGTCGGAGGAGCAGTTCCGCCAGCGCTCGCCCCTGGGTGGAGAT  2414

     201 ValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGl  217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2415 GTGCTGCACGCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAAAGA  2464

     217 uArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluS  234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2465 GCGGACTTCACCTGGGGCGATTCACTACTGTGCCTCCACCAGTGAGGAGA  2514

     234 erTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHis  250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2515 GCTGGACCGACAGCGAGGTGGACTCATCATGCTCCGGGCAGCCCATCCAC  2564

     251 LeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrSlyAr  267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2565 CTGTGGCAGTTCCTCAAGGAGTTGCTACTCAAGCCCCACAGCTATGGCCG  2614

     267 gPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspS  284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2615 CTTCATTAGGTGGCTCAACAAGGAGAAGGGCATCTTCAAAATTGAGGACT  2664

     284 erAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMet  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2665 CAGCCCAGGTGGCCGGCTGTGGGGTATCCGCAAGAACCGTCCCGCCATG  2714

     301 AsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIl  317
         ||||||||||||||||||| ||||||||||||||||||||||||||||||
    2715 AACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAAGGGCAT  2764

     317 eIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheValHisP  334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2765 CATCCGGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACC  2814

     334 roIle 335
         |||||
    2815 CCATC 2819
```

seq_comment /version="1" division="patent" molecule="DNA" EMBL AAF21828

```
seq_documentation_block:
ID   AAF21828 standard; DNA; 1887 BP.
XX
AC   AAF21828;
XX
DT   27-MAR-2001 (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 215.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055174-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA;  Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   P-PSDB; AAB58925.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 1; Page 642-643; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
```

```
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ   Sequence 1087 BP; 215 A; 447 C; 339 G; 182 T; 4 other;


alignment scores:
            Quality: 1560.00              Length:     294
              Ratio:   5.361               Gaps:       0
 Percent Similarity:  99.317   Percent Identity: 99.317

alignment block:
 US-09-126-945B-2 x AAF21828   ..

 Align seq 1/1  to: AAF21828  from: 1  to: 1087

       43 TrpSerProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTy 59
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
        3 TGGAGTCCCAGTCCACCCGCCAGCGCCGAGCAGGGCCTGTCCGCCTTCTA 52

       59 rLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaL 76
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
       53 CCTCTCCTACTTTGACATGCTGTACCCTGAGGACAGCAGCTGGGCAGCCA 102

       76 ysAlaProGlyAlaSerSerArgGluGluProProGluGluProGlnGln 92
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      103 AGCCCCTGGGGCAGCAGTCGGGAGGAGCCACCTGAGGAGCCTGAGCAG 152

       93 CysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValPr 109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      153 TGCCCGGTCATTGACAGCCAAGCCCCAGCGGGCAGCCTGGACTTGGTGCC 202

      109 cGlyGlyLeuThrLeuGluGluHisSerLeuGlnGlnValGlnSerMetV 126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      203 CGGCGGGCTGACCTTGGAGGAGCACTCGCTGCAGCAGGTGCAGTCCATGG 252

      126 alValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsn 142
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      253 TGGTGGGCGAAGTGCTCAAGGACATCGAGACGGCCTGCAAGCTGCTCAAC 302

      143 IleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLe 159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      303 ATCACGGCAGATCCCATGGACTGGAGCCCCAGCAATGTGCAGAAGTGGCT 352

      159 uLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheC 176
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      353 CCTGTGGACAGAGCACCAATACGGGCTGCCCCCCATGGGCAAGGCCTTCC 402

      176 InGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGluPheArg 192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      403 AGGAGCTGGCGGGCAAGGAGCTGTGCGCCATGTCGGAGGAGCAGTTCCGC 452

      193 GlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleIn 209
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      453 CAGCGCTCGCCCCTGGGTGGGGATGTGCTGCACGCCCACCTGGACATCTG 502

      209 pLysSerAlaAlaIrpMetLysGluArgThrSerProGlyAlaIleHisI 226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      503 GAAGTCAGCGGCTTGGATGAAAGAGCGGACTTCACCTGGGGCGATTCACT 552

      226 yrCysAlaSerThrSerGluGluSerTrpThrAspSerGluValAspSer 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      553 ACTGTGCCTCGACCAGTGAGGAGAGCTGGACCGACAGCGAGGTGGACTCA 602

      243 SerCysSerGlyGluProIleHisLeuTrpGlnPheLeuLysGluLeuLe 259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      603 TCAGCCTCCGGCAGCCCATCCACCTGTGGCAGTTCCTCAAGGAGTTGCT 652

      259 uLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGlnL 276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      653 ACTCAAGCCCCACAGCTATGGCCGCTTCATTAGGTGGCTCAACAAGGAGA 702

      276 ysGlyIlePheLysIleGlnAspSerAlaGlnValAlaArgLeuTrpGly 292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      703 AGGCCATCTTCAAAATTGAGGCACTCAGCCCACGTGCCCCGCCTCTGGGC 752

      293 IleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIl 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      753 ATCCGCAACAACCGCCCCGCCATGAACTACGACAAGCTGAGCCGCTCCAT 802

      309 eArgGluTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnA 326
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      803 CCGSCAGTATTACAAGAAGGGCATCATCCGGAAGCCAGACATCTGCCAGC 852

      326 rgLeuValTyrGluPheValHisProIle 335
          ||||||||||||||||||||||||||||
      853 GSCTCGTCTACGAGTTCGTGCACCCCATC 881

Seq name: ???? ?c?i ?ta?Genes pTeces pcDNA2000.2:A1:AAA06617

Seq documentation block:
ID   AAA06617 standard; cDNA; 278 BP.
XX
AC   AAA06617;
XX
DT   13-JUN-2000 (first entry)
XX
DE   Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:498.
XX
KW   Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW   immunogenic; cytostatic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200004149-A2.
XX
PD   27-JAN-2000.
XX
PF   14-JUL-1999;   99WO-US15838.
XX
PR   14-JUL-1998;   98US-0115454.
PR   14-JUL-1998;   98US-0116134.
PR   23-SEP-1998;   98US-0159812.
PR   23-SEP-1998;   98US-0159822.
PR   15-JAN-1999;   99US-0232149.
PR   15-JAN-1999;   99US-0232880.
PR   09-APR-1999;   99US-0288946.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillon DC, Barlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR   WPI; 2000-171268/15.
XX
PT   New polypeptide useful for treating and diagnosing prostate cancer
PT   comprises an immunogenic portion of prostate tumor protein
XX
PS   Claim 50; Page 239; 263pp; English.
XX
CC   The present invention describes isolated polypeptides, comprising an
CC   immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC   and polynucleotides encoding them have cytostatic activity and can be
CC   used in vaccines and in gene therapy. The polypeptides and
CC   polynucleotides encoding them, antigen presenting cells which express
CC   the polypeptides, antibodies against the polypeptides and vaccines
CC   comprising them can be used for inhibiting the development of prostate
CC   cancer in a patient. The polypeptides can be used to generate antibodies
CC   or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC   the polynucleotides encoding the polypeptides can be used as a probe or
CC   to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC   AAY82000 to AAY82020 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 278 BP; 56 A; 85 C; 87 G; 49 T; 1 other;
```

```
alignment_scores:
            Quality:  478.00              Length:     88
              Ratio:  5.494                 Gaps:      0
 Percent Similarity:  98.864   Percent Identity: 97.727

alignment_block:
 US-09-126-945B-2 x AAA06617   ..

 Align seq 1/1  to. AAA06617  from. 1  to. 276

      189 GluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHi 205
          :::|||||||||||||||||||||||||||||||||||||||||||||||
           14 CAGCAGTTCCGCCAGCGCTCGCCCCTGGGTGGGGATGTGCTGCACGCCCA 63

      205 sLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerPro 222
          |||||||||||||||||||||||||||||||||||||||||||||||||
           64 CCTGGACATCTGGAAGTCAGCGGCCTGGATGAAAGAGCGGACTTCACCCG 113

      222 lyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSer 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      114 GGGCGATTCACTACTGTGCCTCGACCAGTGAGGAGAGCTGGACCGACAGC 163

      239 GluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLe 255
          ||||.|||||||||||||||||||||||||||||||||||||||||||||
      164 GAGGTGGACTCATCATGCTCCGGGCAGCCCATCCACCTGTGGCAGTTCCT 213

      255 uLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpL 272
          |||||||||||||||||||||||||||||||||||||||||||   |||
      214 CAAAGGAGTTGCTACTCAAGCCCCACAGCTATGGCCGCTTCATTAGGTGGC 263

      272 euAsnLysGluLys 276
          |||||||||||||
      264 TCAACAAGGAGAAG 277

seq_name  /SIS8/gcgdata/genseq1/ProtSeqDNA2000.DAT AAZ94394

seq_documentation_block:
ID   AAZ94394 standard; cDNA; 898 BP.
XX
AC   AAZ94394;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Human ELF5 cDNA.
XX
KW   ELF5; human; ETS protein; transcription factor; cell proliferation;
KW   cell differentiation; therapy; diagnosis; cancer; tumour;
KW   autoimmune disease; chromosome 11p13-15; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             110..877
FT                   /*tag= a
XX
PN   WO200012695-A1.
XX
PD   09-MAR-2000.
XX
PF   26-AUG-1999;   99WO-AU000691.
XX
PR   27-AUG-1998;   98AU-0005512.
PR   30-SEP-1998;   98AU-0006252.
XX
PA   (MONU ) UNIV MONASH.
XX
PI   Kola I,  Zhou J;
XX
DR   WPI; 2000-270820/24.
DR   P-PSDB; AAY79288.
XX
PI   ELF5 protein, useful for diagnosis, antibody generation and as
PI   screening tool for agents capable of modulating transcriptional events
PI   during cellular functioning such as in tumorigenesis, comprises an ETS
PI   domain  -
XX
PS   Claim 3; Page 86-88; 114pp; English.
XX
CC   The present sequence of that of cDNA encoding novel human ELF5 (see
CC   AAY79288), defined as belonging to the ETS family of transcription
CC   factors due to its expression of an ETS domain which recognises and
CC   binds the purine rich GGA(A/T) core motifs.  ELF5 partial cDNA was
CC   initially isolated from a human lung cDNA library following
CC   screening with a murine ELF5 probe.  The full length coding
CC   sequence of human ELF5 was then obtained by RT-PCR and RACE
CC   using human placental RNA.  The ELF5 gene was mapped to human
CC   chromosome 11p13-15.  ELF5 expression was not detectable in 20/20
CC   human breast carcinomas examined.  The identification of human ELF5
CC   permits the identification and rational design of a range of
CC   products for use in therapy, diagnosis and antibody generation
CC   involving, e.g., regulation of cellular functional activity such as
CC   cellular differentiation or cellular proliferation.  These
CC   therapeutic molecules may also act as either agonists or
CC   antagonists of ELF5 function and will be useful in cancer and
CC   autoimmune disease therapy.
XX
SQ   Sequence 898 BP; 244 A; 221 C; 219 G; 212 T; 0 other;


alignment_scores:
            Quality:  300.00              Length:    262
              Ratio:  2.013                 Gaps:      6
 Percent Similarity:  56.870   Percent Identity: 30.153

alignment_block:
 US-09-126-945B-2 x AAZ94394   ..

 Align seq 1/1  to.  AAZ94394  from. 1  to.  898

      102 AlaGlySerLeuAspLeuValProGlyGlyLeuThrLeu..... 114
          |||:::|||||||||||||:::   ::::::|||||||:::
           71 GCCTCCTCTTTGGACCTAGCCACCGCTGCCCTCACGGTAATGTTGGACTC 120

      115 .GluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluValL 131
               ||||||      ::::::              ::::
      121 GGTGACACACAGCACCTTCCTGCCTAATGCATCCTTCTGCGATCCCCTGA 170

      131 euLys.....................................Asp 133
          :::::                                    :::
      171 TGTCGTGGACTGATCTGTTCAGCAATGAAGAGTACTACCCTGCCTTTGAG 220

      134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIr 150
          :::||||||||              ::::::   |||      ||
      221 CATCAGACAGCCCTCTCACTCATACTCGACATCAGTCCACCCTGAATACTG 270

      150 pSerProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGluTyrA 167
          |:::      :::|||   :::|||||   :::          ||||||:
      271 GACTAAACGCCATCTGTGGGAGTGCCTCCAGTTCTGCTGCGACCAGTACA 320

      167 rgLeuProProMetGlyLysAlaPhe......GlnGluLeuAlaGlyLys 181
          :::|||           :::|||          ::::::::::|||
      321 AGTTGGACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGGCCTG 370

      182 GluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGl 198
          :::|||||:::||||:::|||||||   :::   :::   |||
      371 CAGCTGTGCAGCATGACACAGGAGCAGTTCGTCGAGGCAGCTGGCCTCTG 420

      198 yGlyAspValLeuHisAlaHisLeu..................... 206
          |||:::   |||:::      |||
      421 GGGCGAGTACCTGTACTTCATCCTCCAGAACATCCGCACACAAGGTTACT 470

      207 ..........AspIleTrpLysSerAlaAlaTrpMetLysGluArgThr 219
                    |||     ::::||   |||   :::|||:::   :::
      471 CCTTTTTTAATGACGCTGAAGAAAGCAAGGCCACCATCAAAGACTATGCT 520
```

```
      220 SerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpTh 236
          :::   :::             |||   :::|||||      :::
      521 GATTCCAAC............TGCTTGAAAACAAGTGGCATCAAAAGTCA 558

      236 rAspSerGluValAspSerSerCysSerGlyGluProIleHisLeuTrpC 253
          |||   :::       |||      |||   |||      ||||||||||
      559 AGACTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGG 608

      253 InPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 269
          ::|||||||||||||||||||||||:::||||||||      ::::::
      609 AATTTCTACGACACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTG 658

      270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 286
          :::|||   ::::::|||:::|||||||||||:::::      |||
      659 GAATGGAAGATAGGAACAAGAATTTTCGGGTGGTTAAATCGCAAGC 708

      286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 303
          :::|||:::::|||||   ||||||:::::      |||:::|||:
      709 CCTGGCAAAGATGTGGGACAAAGGAAGAAAAATGACAGAATGACATATG 758

      303 spLysLeuSerArgSerIleArgGlaTyrTyrLysLysGlyIleIleArg 319
          ::|||||||||||||:::::|||   ||||||||   ||||||:::::
      759 AAAAGTTCAGCAGAGCTCTGAGATACTACTATAAAACAGGAATTTTGGAG 808

      320 LysProAspIleSerGluArgLeuValTyrGlnPhe 334
          :::   |||       :::|||||||||||||:|||
      809 CGGGTTGAC.....CGAAGGTTACTGTACAAATTT 838

Seq name: gi|11038|gb|AAA52484.1|AF000230_1 Seq_pse:Seq_NAT:Pct:AT 66.1%

Seq documentation block:
ID   AAZ18360 standard; cDNA; 2280 BP.
XX
AC   AAZ18360;
XX
DT   19-OCT-1999  (first entry)
XX
DE   Human ASTHL1 alt2 cDNA sequence.
XX
KW   ASTHL1; asthma; human; chromosome 11p; ASTHL1; ASTHL2; genetic locus;
KW   therapeutic; immunogen; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             71..838
FT                   /*tag= a
FT                   /product= "ASTHL1 alt2 protein"
XX
PN   WO9937809-A1.
XX
PD   29-JUL-1999.
XX
PF   21-JAN-1998;   98WO-US01260.
XX
PR   21-JAN-1998;   98WO-US01260.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Brooks-Wilson AR;  Buckler A;  Cardon L;  Carey AH;
PI   Galvin M;  Miller A;  North M;
XX
DR   WPI; 1999-479058/40.
DR   P-PSDB; AAY81418.
XX
PT   Mammalian asthma related genes, useful for diagnosis of a
PT   predisposition to development of asthma
XX
PS   Claim 15; Page 106; 195pp; English.
XX
CC   The invention identifies a genetic locus ASTHL, associated with asthma,
CC   mapped to human chromosome 11p. ASTHL1 and ASTHL2 are genes present
CC   within the locus, located close to each other on human chromosome 11p,
CC   and have similar patterns of expression, and common sequence motifs. The
CC   ASTHL genes and fragments, encoded protein, genomic regulatory regions
CC   and anti-ASTHL antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for the modulation of gene
CC   activity in vivo for prophylactic and therapeutic purposes. the ASTHL
CC   protein is useful as an immunogen to raise specific antibodies, in drug
CC   screening for compositions that mimic or modulate ASTHL activity or
CC   expression, including altered forms of ASTHL protein, and as a
CC   therapeutic. The present sequence represents a human ASTHL1 alt2
CC   cDNA sequence.
XX
SQ   Sequence 2280 BP; 676 A; 481 C; 472 G; 651 T; 0 other;


alignment scores:
            Quality:  800.00              Length:    262
              Ratio:  2.014                 Gaps:      6
 Percent Similarity:  56.870   Percent Identity:  40.154

alignment block:
 US-09-126-945B-2 x AAZ18360   ..

 Alignment 1/1   1-- to AAZ18360   from: 1   to: 2280

      102 AlaGlySerLeuAspLeuValProGlyGlyLeuThrLeu..... 114
          ||||:|||||||||||||||:::   ::::::||||||:::
       82 GCCTTCTCTTTGGACCTAGCCACTCGTGCCCTCACGGTAATGTTGGACTC 81

      .15 .GLy.Ala.Ser.Leu...Pro.Val.Leu.Met.Val.Ala.Gly.Leu.Val.L 131
                ||||||              ::::::          ::::
       82 GGTGACACACAGCACCTTCCTGCCTAATGCATCCTTCTGCGATCCCCTGA 131

      131 enLys.....................................Asp 133
          :::::                                   :::
      132 TGTCGTGGACTGATCTGTTCAGCAATGAAGAGTACTACCCTGCCTTGAG 181

      134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTr 150
          :::|||||||||              ::::::   |||      ||
      182 CATCAGACAGCCTGTGACTCATACTGGACATCAGTCCACCCTGAATACTG 231

      150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrA 167
          |:::      :::|||   :::||||||   :::          ||||||
      232 GACTAAGCGCCATGTGTGGGAGTGGCTCCAGTTCTGCTGCTGCACCAGTACA 281

      167 rqLeuProProMetGlyLysAlaPhe......GlnGluLeuAlaGlyLys 181
          ::|||          :::|||          ::::::::::|||
      282 AGTTGGACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGGCCTG 331

      182 GluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGl 198
          :::|||||||:::|||:::::|||:::|||   :::   :::   |||
      332 CAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAGCTGGCCTCTG 381

      198 yGlyAspValLeuHisAlaHisLeu..... 206
          |||:::   |||:::      |||
      382 CGGCGAGTACCTGTACTTCATCCTCCAGAACATCCGCACACAACGTTACT 431

      207 .....AspIleTrpLysSerAlaAlaTrpMetLysGluArgThr 219
                |||      :::|||   |||   :::|||:::   :::
      432 CCTTTTAATGACGTGAAGAAAGCAAGGCCACCATCAAAGACTATGCT 481

      220 SerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpTh 236
          :::   :::             |||   :::|||||      :::
      482 GATTCCAAC............TGCTTGAAAACAAGTGGCATCAAAAGTCA 519

      236 rAspSerGluValAspSerSerCysSerGlyGluProIleHisLeuTrpC 253
          |||   :::       |||      |||   |||      ||||||||||
      520 AGACTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGG 569

      253 InPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 269
          ::|||||||||||||||||||||||:::||||||||      ::::::
      570 AATTTCTACGACACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTG 619
```

```
      270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 296
          :::|||   ::::::|||:::|||||||||||:::::      |||
      629 GAATGGGAAGATAGGGAACAAGGAATTTTTCGGGTGGTTAAATCGGAAGG 680

      286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 312
          :::|||:::::::|||||   ||||||:::::      |||:::|||:
      670 CCTGGCAAAGATGTGGGCACAAAGGAAGAAAAATGACAGAATGACGTATG 719

      303 spLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArg 329
          :::|||||||||||:::::|||   |||||||||   ||||||:::::
      720 AAAAGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGGAATTTTGGAG 769

      326 LysProAspIleSerGluArgLeuValTyrGlnPhe 351
          :::   |||      :::|||||||||||||:|||
      770 CGGGTTGAC......CGAAGGTTAGTGTACAAATTT 799

seq_name: /S1.s82/qpalm2/2/2.seq_homosapiens_NA2000_2_AT_AAAP.rpt

seq_documentation_block:
ID   AAA80258 standard; cDNA; 2280 BP.
XX
AC   AAA80258;
XX
DT   22-NOV-2000 (first entry)
XX
DE   Human ASTH1I cDNA, splice variant alt2.
XX
KW   ASTH1 locus; ASTH1I; human; chromosome 11p; asthma;
KW   bronchial hyperreactivity; ets family; transcription factor;
KW   splice variant; genetic predisposition; polymorphism; antibody;
KW   drug screening; prophylaxis; therapy; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             70..849
FT                   /*tag= a
FT                   /product= "Human ASTH1I isoform alt2"
XX
PN   US6087485-A.
XX
PD   11-JUL-2000.
XX
PF   21-JAN-1998;    98US-0009913.
XX
PR   21-JAN-1997;    97US-0035664.
PR   01-JUL-1997;    97US-0051432.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M;  Miller A;  North M;  Cardon I;  Buckler A;
PI   Brooks-Wilson AR;  Carey AH;
XX
DR   WPI; 2000-505109/45.
DR   P-PSDB; AAB11970.
XX
PT   New nucleic acids other than naturally occurring chromosomes encoding
PT   ASTH1 protein, for e.g. screening compositions that modulate expression
PT   or function of ASTH1 proteins or as diagnostics for genetic
PT   predisposition to asthma -
XX
PS   Examples; Column 133-134; 131pp; English.
XX
CC   The invention relates to the ASTH1 locus on the short arm of human
CC   chromosome (11p). This locus comprises the ASTH1I and ASTH1J genes,
CC   which are associated with a genetic predisposition to asthma and
CC   bronchial hyperreactivity. The ASTH1I and ASTH1J genes are oriented in
CC   opposite directions with the ASTH1 locus, and have similar patterns of
CC   expression and common sequence motifs. They are both expressed in
CC   trachea, lung and several other tissues. ASTH1I and ASTH1J are novel
CC   members of the ets family of transcription factors, which have been
CC   implicated in the activation of a variety of genes including the TCRa
CC   gene and cytokine genes known to be important in the activation of
CC   asthma. Both ASTH1I and ASTH1J mRNAs are alternatively spliced.
CC   Alternative splicing of transcripts has no effect on the open reading
CC   frame of ASTH1J, as the exons involved are all 5' to the start codon in
CC   exon b. In contrast, alternative splicing of ASTH1I transcripts results
CC   in 3 different ASTH1I isoforms. The invention also encompasses mouse
CC   asth1j protein. The ASTH1 nucleic acids are useful as diagnostics to
CC   identify a hereditary predisposition to asthma, as probes for identifying
CC   ASTH1 related genes, for identifying expression of the gene in a
CC   biological specimen, and for generating genetically modified non human
CC   animals or site specific gene modifications in cell lines. The encoded
CC   ASTH1 proteins are useful as immunogens to raise specific antibodies; in
CC   drug screening for compositions that mimic or modulate activity or
CC   expression of ASTH1I and/or ASTH1J (including altered forms of these
CC   proteins); and as a therapeutic. The ASTH1 genes or fragments thereof,
CC   encoded proteins, ASTH1 genomic regulatory regions, and anti-ASTH1I and
CC   anti-ASTH1J antibodies are useful in the identification of individuals
CC   predisposed to development of asthma and in modulating [illegible]
CC   in vivo for prophylactic and therapeutic purposes. The intact ASTH1I or
CC   ASTH1J proteins or active fragments thereof may be used to modulate or
CC   reduce bronchial hyperreactivity. Sequences AAA80257 A80259 represent
CC   three alternatively spliced human ASTH1I cDNAs which encode three
CC   different isoforms of ASTH1I (AAB11969-B11971).
XX
SQ   Sequence 2280 BP; 676 A; 481 C; 472 G; 651 T; 0 other;


alignment_scores:
            Quality:  200.00              Length:    262
              Ratio:  2.013                 Gaps:      6
 Percent Similarity:  54.979   Percent Identity: 30.153

alignment_block:
 US-09-126-945B-2 x AAA80258   ..

 Align seq 1/1  to: AAA80258  from: 1  to: 2280

      102 AlaGlySerLeuAspLeuValProGlyGlyLeuThrLeu..... 114
          |||:::|||||||||||||:::   ::::::::::::::
       32 GCCTCCTCTTTGGACCTAGCCACCGCTGCCCTCACGGTAATGTTGGACTC 81

      115 .GluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluValL 141
             ||||||             ::::::             ::::
       82 GGTGACACACAGCACCTTCCTGCCTAATGCATCCTTCTGCGATCCCCTGA 131

      131 euLys.....Asp 133
          :::::       :::
      132 TGTCGTGGACTGATCTGTTCAGCAATGAAGAGTACTACCCTGCCTTTGAG 181

      134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTr 150
          :::|||||||||||             ::::::   |||      ||
      182 CATCACACAGCCTGTGACTCATACTGGACATCAGTCCACCCTGAATACTG 231

      150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrA 167
          |:::      :::|||   ::|||||||   :::          ||||||:
      232 GACTAAGCGCCATGTGTGGCAGTGGCTCCAGTTCTGCTCCGACCAGTACA 281

      167 rgLeuProProMetGlyLysAlaPhe......GlnGluLeuAlaGlyLys 181
          ..|||               :::|||         ::::::::::|||
      282 AGTTGCACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGGCCTG 331

      182 GluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGl 198
          :::|||||:::|||:::::|||:::|||   :::   :::   |||
      332 CAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAGCTGGCCTCTG 381

      198 yGlyAspValLeuHisAlaHisLeu..... 206
          |||:::   |||:::   |||
      382 CGGCGAGTACCTGTACTTCATCTCCAGAACATCCGCACACAAGGTTACT 431

      207 .....AspIleTrpLysSerAlaAlaTrpMetLysGluArgThr 219
                |||      :::|||   |||   :::|||:::   :::
      432 CCTTTTTTAATGACGCTGAAGAAACCAAGCCCACCATCAAAGACTATGCT 481

      220 SerProGlyAlaIleHisTrpLysAlaSerThrSerGlnGluSerTrpTh 246
```

```
      482 GATTCCAAC.............TGCTTGAAAACAAGTGGCATCAAAAGTCA 519

      246 rAspSerGluValAspSerSerCysSerGlyGluProIleHisLeuArgC 258
          |||   :::     |||     |||   |||       ||||||||||
      520 AGACTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGG 569

      259 InPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 269
          ::|||||||||||||||||||||::|||||||       ::::::
      570 AATTTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTCTGGCATTCTG 619

      270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 286
          :::|||   ::::::|||:::|||||||||:::::       |||
      620 GAATGGAAGATACGGAACAAGGAATTTTCGGGTGGTTAAATCGGAAGC 669

      286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 303
          :::|||:::::|||||   ||||||:::::      |||:::|||:
      670 CCTGGCAAAGATGTGGGACAAAGGAAGAAAATGACAGAATGACGTATG 719

      303 spLysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleIleArg 319
          ::|||||||||||||:::::|||   ||||||||   ||||||:::::
      720 AAAAGTTGAGCAGAGCCTGAGATACTACTATAAAACAGGAATTTGGAG 769

      320 LysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          :::   |||       :::|||||||||||||:::|||
      770 CGGCTTGAC.....CGAAGCTTAGTGTACAAATTT 799

Seq_name: /SIDS82/prodat2/geneSeq/geneseqp/NA1999.DAT:AAZ18361

Seq_documentation block:
ID   AAZ18361 standard; cDNA; 2498 BP.
XX
AC   AAZ18361;
XX
DT   19-OCT-1999  (first entry)
XX
DE   Human ASTHI1 alt3 cDNA sequence.
XX
KW   ASTHI; asthma; human; chromosome 11p; ASTHI1; ASTHIJ; genetic locus;
KW   therapeutic; immunogen; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             561..1055
FT                   /*tag= a
FT                   /product= "ASTHI1 alt3 protein"
XX
PN   WO9937809-A1.
XX
PD   29-JUL-1999.
XX
PF   21-JAN-1998;   98WO-US01260.
XX
PR   21-JAN-1998;   98WO-US01260.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Brooks Wilson AR;  Buckler A;  Cardon L;  Carey AH;
PI   Galvin M;  Miller A;  North M;
XX
DR   WPI; 1999-478058/40.
DR   P-PSDB; AAY41419.
XX
PT   Mammalian asthma related genes, useful for diagnosis of a
PT   predisposition to development of asthma
XX
PS   Claim 15; Page 108; 195pp; English.
XX
CC   The invention identifies a genetic locus ASTHI, associated with asthma,
CC   mapped to human chromosome 11p. ASTHI1 and ASTHIJ are genes present
CC   within the locus, located close to each other on human chromosome 11p,
CC   and have similar patterns of expression, and common sequence motifs. The
CC   ASTHI genes and fragments, encoded protein, genomic regulatory regions
CC   and anti-ASTHI antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for the modulation of gene
CC   activity in vivo for prophylactic and therapeutic purposes. The ASTHI
CC   protein is useful as an immunogen to raise specific antibodies, in drug
CC   screening for compositions that mimic or modulate ASTHI activity or
CC   expression, including altered forms of ASTHI protein, and as a
CC   therapeutic. The present sequence represents a human ASTHI1 alt3
CC   cDNA sequence.
XX
SQ   Sequence 2498 BP; 706 A; 554 C; 554 G; 685 T; 0 other;


Alignment scores:
            Quality:  299.50              Length:     350
              Ratio:  1.576                 Gaps:      13
 Percent Similarity:  54.286   Percent Identity: 29.429

Alignment block:
 US-09-126-945B-2 x AAZ18361   ..

 Align seq 1/1  to: AAZ18361  from: 1  to: 2498

        4 SerAlaSerProGlyLeuLeuSerSerValSerProSerHisLeuLeuLeuP 19
          :::|||||||   ||||||   :::::::::   ||||||      :::|
       85 GCGGGCTCTGGAGGCCTGCCATGGCACCCAGAGCAGCCATC.....TTTC 128

       20 rssrsAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAla 35
          ||||:::::::                  |||      ||||||
      129 CTCCCAGTTCT.....................GGGGCTTTGGAAGGAGCT 157

       36 ValGlyLeuGluArgArg......AspTrpSerProSerProProAlaT 50
             |||    ||||||||:::      :::   |||   ||||||||||::::
      158 TGCGGATGAGGAGAGGGAGCCTCCGCAGGGCTCTGGCTCCCCTCCAGGGG 207

       50 hrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeu 66
          ::                  ::|||                    |||
      208 CCGAGGCCGCACACAAAGCCGCT.....................CTG 233

       67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 83
          :::|||:::   :::   |||   :::          |||
      234 TGGCCCAATTACACCTACTGGATAGGATTGTTGAGGGGA........... 272

       83 qGlnGlnProProGlnGluProGluGlnCysProValIleAspSerGlnA 100
                 |||||||||   |||          :::   ::::::::::|
      273 .........CCTGAGAAACCTGAGACGACAAGAACGCGTAGCGCCTCGG 312

      100 LaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGlu... 115
          ||      |||         ||||||     :::::|   ||||::
      313 CTGGCTGAGGGTGCTGAGCCCTCGTGTTGTGTTCTCTCCAGCTTTCCCCG 362

      116 ......GluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
                ::|||||||   :::::::::   ::::::
      363 TGCCTCAGCCACTCTTCACGTTCCATCTGTGCTCTG.............. 398

      130 IleuLysAspIleGlaThrAlaCysLysLeuLeuAsnIleThrAlaAspP 147
                              ||||||                 ::::::   |
      399 .........TGCTGACCGGCTGTGACTCATACTGGACATCACGTCCACC 438

      147 roMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGlu 163
          ||      ||::::      :::|||   :::|||||   :::
      439 CTGAATACTGGACTAAGCGCCATGTGTGGGAGTGGCTCCACTTCTGCTGC 488

      164 HisGlnTyrArgLeuProProMetGlyLysAlaPhe......GlnGluLe 178
             ||||||:::|||               ::|||          :::::
      489 GACCACGTACAAGTTGGACACCAATTGCATCTCCTTCTGCAACTTCAACAT 538

      178 uAlaGlyLysSiaLeuCysAlaMetSerGluGluGlnPheArgGluArgS 195
          ::::|||   :::|||||||:::|||||:::::|||:::|||   :::   :
      539 CAGTGGCCTGCAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAG 588

      195 erProLeuGlyGlyAspValLeuHisAlaHisLeu.............. 206
          ::   |||   ||:::   ||:::      |||
```

```
 589 CTGGCCTCTGCGGCGAGTACCTGTACTTCATCCTCCAGAACATCCGCACA 638

 207 .....AspIleTrpLysSerAlaAlaTrpMetLy 216

 639 CAAGGTTACTCCTTTTTTAATGACGCTGAAGAAAGCAAGGCCACCATCAA 688

 216 sGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluG 233

 689 AGACTATGCTGATTCCAAC.....TGCTTGAAAACAAGTGGCA 726

 233 luSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIle 249

 727 TCAAAAGTCAAGACTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCT 776

 250 HisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGl 266

 777 CATCTATGGGAATTTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTG 826

 266 yArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluA 283

 827 TGGCATTCTGGAATGGGAAGATAGGGAACAAGGAATTTTCGGGTGGTTA 876

 283 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299

 877 AATCGGAAGCCCTGGCAAAGATGTGGGGACAAAGGAAGAAAAATGACAGA 926

 300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316

 927 ATGACGTATGAAAAGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGG 976

 316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331

 977 AATTTTGGAGCGGGTTGAC.....CGAAGGTTAGTGTACAAATTT 1016
```

seq_name: /SIDS8/qcqdata/geneseq/geneseqn/NA2000_DAT:AAA80259

```
seq_documentation_block:
ID   AAA80259 standard; cDNA; 2498 BP.
XX
AC   AAA80259;
XX
DT   22-NOV-2000  (first entry)
XX
DE   Human ASTH1I cDNA, splice variant alt3.
XX
KW   ASTH1 locus; ASTH1I; human; chromosome 11p; asthma;
KW   bronchial hyperreactivity; ets family; transcription factor;
KW   splice variant; genetic predisposition; polymorphism; antibody;
KW   drug screening; prophylaxis; therapy; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             561..1055
FT                   /*tag= a
FT                   /product= "Human ASTH1I isoform alt3"
FT
XX
PN   US6087485-A.
XX
PD   11-JUL-2000.
XX
PF   21-JAN-1998;    98US-0009913.
XX
PR   21-JAN-1997;    97US-0035663.
PR   01-JUL-1997;    97US-0051432.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M, Miller A, North M, Cardon L, Buckler A;
PI   Brooks-Wilson AR, Carey AH;
XX
DR   WPI; 2000-505109/45.
DR   P-PSDB; AAB11971.
XX
PT   New nucleic acids other than naturally occurring chromosomes encoding
PT   ASTH1 protein, for e.g. screening compositions that modulate expression
PT   or function of ASTH1 proteins or as diagnostics for genetic
PT   predisposition to asthma -
XX
PS   Examples; Column 135-138; 131pp; English.
XX
CC   The invention relates to the ASTH1 locus on the short arm of human
CC   chromosome (11p). This locus comprises the ASTH1I and ASTH1J genes,
CC   which are associated with a genetic predisposition to asthma and
CC   bronchial hyperreactivity. The ASTH1I and ASTH1J genes are oriented in
CC   opposite directions with the ASTH1 locus, and have similar patterns of
CC   expression and common sequence motifs. They are both expressed in
CC   trachea, lung and several other tissues. ASTH1I and ASTH1J are novel
CC   members of the ets family of transcription factors, which have been
CC   implicated in the activation of a variety of genes including the TCRa
CC   gene and cytokine genes known to be important in the aetiology of
CC   asthma. Both ASTH1I and ASTH1J mRNAs are alternatively spliced.
CC   Alternative splicing of transcripts has no effect on the open reading
CC   frame of ASTH1J, as the exons involved are all 5' to the start codon in
CC   exon b. In contrast, alternative splicing of ASTH1I transcripts results
CC   in 3 different ASTH1I isoforms. The invention also encompasses mouse
CC   asth1j protein. The ASTH1 nucleic acids are useful as diagnostics to
CC   identify a hereditary predisposition to asthma, as probes for identifying
CC   ASTH1 related genes, for identifying expression of the gene in a
CC   biological specimen, and for generating genetically modified non-human
CC   animals or site specific gene modifications in cell lines. The encoded
CC   ASTH1 proteins are useful as immunogens to raise specific antibodies; in
CC   drug screening for compositions that mimic or modulate activity or
CC   expression of ASTH1I and/or ASTH1J (including altered forms of these
CC   proteins); and as a therapeutic. The ASTH1 genes or fragments thereof,
CC   encoded proteins, ASTH1 genomic regulatory regions, and anti-ASTH1I and
CC   anti-ASTH1J antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for modulation of gene activity
CC   in vivo for prophylactic and therapeutic purposes. The intact ASTH1I or
CC   ASTH1J proteins or active fragments thereof may be used to modulate or
CC   reduce bronchial hyperreactivity. Sequences AAA80257-A80259 represent
CC   three alternatively spliced human ASTH1I cDNAs which encode three
CC   different isoforms of ASTH1I (AAB11969-B11971).
XX
SQ   Sequence 2498 BP; 706 A; 553 C; 554 G; 685 T; 0 other;
```

```
alignment_scores:
            Quality:  299.50          Length:    350
              Ratio:  1.576             Gaps:     13
 Percent Similarity:  54.286  Percent Identity: 29.429

alignment_block:
 US-09-126-945B-2 x AAA80259   ..

 Align seq 1/1  to: AAA80259  from: 1  to: 2498

       3 SerAlaSerProGlyLeu.SerSerValSerProSerHisLeuLeuLeuP 19

      85 GCGGCCTCTGGAGGCCTGCCATGGCACCCAGAGCAGCCAT.....TTTC 128

      19 roProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAla 35

     129 CTCCCAGTTCT.....GGGGCTTTGGAAGGAGCT 157

      36 ValGly.LeuGluArgArg.....AspTrpSerProSerProProAlaP 50

     158 TGCGGATGAGGAGAGGGAGCCTCCGCAGGGCTCTGGCTCCCCTCCAGGGG 207

      50 hrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeu 66

     208 CCGAGGCCGCACACAAAGCCGCT.....CTG 234

      67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 83

     234 TGGCCCAATTACACCTACTGGATAGGATTGTTGAGGGGA..... 272
```

```
 88 qGlnGluProProGluGlnProGluGlnCysProValIleAspSerGluA 100
                |||||||:::  |||          :::  :::::::::|
273 .........CCTGACAAACTTCACACGACAAGAACGCGTACGCGCTCTCC 312

100 laProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGlu... 115
    ||      |||         ||||||      :::|||  |||:::
313 CTGGCTCAGGGTGCTGAGCCCTCGTGTTGTCTTCTCCAGCTTTCCCCG 362

116 ......GluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 139
          :::||||||  :::::::::  ::::::
363 TGCCTCAGCCACTCTTCACGTTCCATCTGTGCTCTG.............. 398

140 lLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspP 147
                  ||||||                  ::::::   |
399 .........TGCTGACCCGCCTGTGACTCATACTGGACATCAGTCCACC 448

147 roMetAspTrpSerProSerAsnValGluLysTrpLeuLeuTrpThrGlu 164
    ||      ||:::      :::|||  :::||||||  :::
449 CTGAATACTGGACTAAGCGCCATGTGTGGGAGTGGCTCCAGTTCTGCTGC 498

164 HisGluTyrArgLeuProProMetGlyLysAlaPhe......GlnGluLe 178
       ||||||:::|||                :::|||       :::::
489 GACCAGTACAAGTTGGACACCAATTGCATCTCCTTCTGCAACTTCAACAT 538

178 uAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGluArgS 195
    :::|||  :::||||||:::|||:::::|||:::|||  :::   :
539 CAGTGGCCTGCAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAG 588

195 erProLeuGlyGlyAspValLeuHisAlaHisLeu.............. 206
    ::  |||  |||:::  |||:::      |||
589 CTGGCCTCTGCGGCGACTACCTGTACTTCATCCTCCAGAACATCGGCACA 638

207 ....................AspIleTrpLysSerAlaAlaTrpMetLy 226
                        |||       :::||  |||  ::::|
639 CAAGGTTACTCCTTTTTTAATGACGCTGAACAAAGCAAGGCCACCATCAA 688

226 sGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGlnL 245
    |:::  ::::::  :::            |||  :::||||||
689 AGACTATCCTGATTCCAAC...........TGCTTCAAAACAAGTGGCA 726

245 luSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIle 264
      :::      |||  :::      |||      |||  |||
727 TCAAAAGTCAAGACTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCT 776

265 HisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGl 284
    ||||||||||:::||::::::::::||||||||||:::||:::::::
777 CATCTATGGGAATTTGTACCACACCTGCTTCTATCTCCTGAAGAAAACTG 826

285 yArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleAla 304
      ::::::::::|||   ::::::|||:::|||||||||:::::
827 TGGCATTCTGGAATGGCAAGATAGGCAACAAGGAATTTTCGGCTGGTTA 876

305 spSerAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 324
      |||      :::|||:::::::|||||   ||||||:::::
877 AATCGGAAGCCCTGGCAAAGATGTGGGGACAAAGGAAGAAAAATGACAGA 926

325 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 344
    |||:::|||:::|||||||||||:::::::|||   ||||||||   ||
927 ATGAGTATGAAAAGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGG 976

345 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 361
    ||||::::::::::   |||      :::|||||||||||:::|||
977 AATTTTGGAGCGGGTTGAC.....CGAAGGTTAGTGTACAAATTT 1019
```

Seq name: [illegible] pg_NA2000.DAT:AAA38442

Seq documentation block:

```
ID   AAA38442 standard; DNA; 2268 BP.
XX
AC   AAA38442;
XX
DT   21 AUG 2000  (first entry)
XX
DE   DNA encoding human Ets-2 transcription factor.
XX
KW   Ets 2; human; transcription factor; chromosome 21q22.3; cancer; invasion;
KW   metastasis; skeletal abnormality; Down's syndrome; expression inhibition;
KW   antisense therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             291..1700
FT                   /*tag= a
FT                   /product  "Human Ets-2"
XX
PN   US6054416-A.
XX
PD   25-APR-2000.
XX
PF   25-JUN-1999;   99US-0344579.
XX
PR   25 JUN 1999;   99US-0344579.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Baker BF, Cowsert LM;
XX
DR   WPI; 2000-338495/29.
DR   P-PSDB; AAY90689.
XX
PT   Antisense compound, 8-30 nucleobases in length, inhibiting the
PT   expression Ets-2 is useful for treating cancer and detecting Ets-2
PT   expression -
XX
PS   Example 13; Columns 41 46; 81pp; English.
XX
CC   This sequence represents DNA encoding the human transcription factor
CC   Ets-2. The Ets domain transcription factors are a family of proteins
CC   which are involved in controlling key cellular events such as
CC   proliferation, differentiation and development. The Ets domain is a
CC   DNA-binding domain shared by all members of this family. Through this
CC   motif, Ets family members bind to the promoter regions of various genes
CC   at a GCA consensus sequence, thereby acting as either repressors or
CC   activators of the gene. All but one Ets family protein bind to DNA as a
CC   monomer. Ets-2 has been implicated in the regulation of cellular
CC   proliferation and differentiation. The Ets-2 gene is located at
CC   chromosome 21q22.3, which is within a region known to undergo
CC   translocations associated with malignancies. Ets-2 has been found to be
CC   upregulated in several cancers, including lymphoblastic leukaemia. It
CC   may also play a role in the cancer phenotype, as it activates the
CC   urokinase plasminogen activator (uPA) promoter and the promoters of
CC   metalloproteinases in response to epidermal growth factor (EGF)
CC   stimulation. High levels of uPA and metalloproteinases are associated
CC   with tumour invasion and metastasis in breast cancers. As the Ets-2 gene
CC   is located on chromosome 21, which is triplicated in Down's syndrome,
CC   it is also thought to be responsible for the skeletal abnormalities
CC   present in this condition. The invention relates to antisense
CC   oligonucleotides targetted to the human Ets-2 gene, which inhibit its
CC   expression. A series of oligonucleotides (AAA38349-A38388) were designed
CC   to target different regions of the human Ets-2 RNA, and were analysed
CC   for their effect on Ets-2 mRNA levels by quantitative real-time PCR. The
CC   antisense oligonucleotides of the invention are useful for the treatment
CC   or prophylaxis of conditions associated with Ets-2 expression,
CC   especially cancer.
XX
SQ   Sequence 2268 BP; 550 A; 642 C; 584 G; 492 T; 0 other;
```

```
alignment_scores:
            Quality:  299.00              Length:     562
              Ratio:  1.484                 Gaps:      18
 Percent Similarity:  39.130   Percent Identity: 22.464

alignment_block:
```

```
US-09-126-945B-2 x AAA38342  ..

Align seq 1/1  to: AAA38342  from: 1  to: 2268

      3 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
        ||:::   |||      |||:::|||   ||||||       ||||||||
      5 TCCTCCCCTCCCCTCCACTCGGCCGTCCCTCCTTCCTCCICCCTCCTCCC 54

     19 oProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaV 36
        ||||     :::   :::         |||:::||||||   |||:::::
     55 TCCTCCTCCCGCTCCTGAA.....    GAGCGCGCCGCGTGGGGCACGG 95

     36 alGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu 52
        ::   |||                        ||||||   |||   |||
     96 CCCGGTTA                      CTTCCTCCAGAGACTGACGAG 124

     53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPhe......AspMetLeuTy 67
           |||:::::              |||                 :::
    125 TGCGGIGIGGCTCCAGCTCAGAGCTCCCGGAGCCGCCCGCCAGCGTCCG 174

     67 rProGluAspSerSer TrpAlaAlaLysAlaPro                78
          |||      |||||| |||   |||   :::|||
    175 GCCTGCCTGATCGTCTCTGGCCGGCGCCCTCGCCCTCGCCCGGCGCGCAC 224

     79 ...GlyAlaSerSerArgGluGluProProGluGluProGluGlnCysPr 94
           :::|||:::::   |||:::|||||:::::|||:::      ||
    225 CGAGCAGCCGCGGGCGCCGAGCAGCCACCGTCCCGAGCAAGCCCCGCCCC 274

     94 oVal.....................................IleAspSerG 99
        |:::                                    :::||||:::
    275 TCCCCGCAGCGGCAGGATGAATGATTTCGGAATCAAGAATATGGAGCAGG 324

     99 lnAlaProAlaGly......                            103
          ||||||:::::
    325 TAGCCCCTGTGGCTAACAGTTACAGAGGGACACTCAAGCGCCAGCCAGCC 374

    103 ....                                       ..... 108

    375 TTTGACACCTTTCATGGGTCCCTGTTTGCTGTTTTTCCTTCTCTAAATGA 424

    104 .......SerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisS 118
              :::|||:::   ||||||   ||||||         :::|||:
    425 AGAGCAAACACTGCAAGAAGTGCCAACAGGCTTGGATTCCATTTCTCATG 474

    118 er                          leuGluGlnValGlnSerMetVal 126
        ::                          |||         ::::::   |||
    475 ACTCCGCCAACTGTGAATTGCCTTTGTTAACCCCGTGCAGCAAGGCTGTG 524

    127 ValGlyGluValLeuLysAspIleGluThrAlaCysLys..... 139
        :::::::::::::||||||       :::::   |||
    525 ATGAGTCAAGCCTTAAAAGCTACCTTCAGTGGCTTCAAAAAGGAACAGGG 574

    140 .LeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValG 156
         |||::::||      :::|||     ||||||   :::::::|||
    575 GCGCCTGGGCATTCCAAAGAACCCCTGGCTGTGGAGTGAGCAACAGGTAT 624

    156 lnLysTrpLeuLeuTrpIhrGluHisGlnTyrArgLeuProProMetGly 172
          :::|||||||||||:::   :::::::   |||      :::::
    625 GCCAGTGGCTTCTCTGGGCCACCAATGAGTTCAGTCTGCTGAACGTGAAT 674

    173 LysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGl 189
                    :::   |||:::   ||||||   ::::::::|||
    675 CTGCAGAGGTTCGGCATGAATGGCCAGATGCTGTGTAACCTTGGCAAGGA 724

    189 uGlnPheArgGlnArgSerPro...LeuGlyGlyAspValLeuHisAlaH 205
        |:::|||   :::   :::|||   :::   ||||||:::|||      |
    725 ACGCTTTCTGGAGCTGCCACCTCACTTTGTGGGTGACATTCTCTGGGAAC 774

    205 isLeuAsp..................................... 207
        ||||||:::
    775 ATCTGGAGCAAATGATCAAAGAAAACCAAGAAAAGACAGAAGATCAATAT 824

    208 ..............IleTrpLysSerAlaAlaTrpMet.......... 215
```

```
                        :::   :::          |||:::
    825 GAAGAAAATTCACACCTCACCTCCGTTCCTCATTGGATTAACAGCAATAC 874

    215 ..................................................... 215

    875 ATTAGGTTTTGGCACAGAGCAGGCGCCCTATGGAATGCAGACACAGAATT 924

    215 ..................................................... 215

    925 ACCCCAAAGCCCGCCTCCTGCACAGCATGTGTCCGGCCTCCACACCCAGC 974

    216 ..................................LysGluArgThrSe 220
                                                ||||:|||   ||
    975 GTACTCAGCTCTGAGCAGGAGTTTCAGATGTTCCCCAAGTCTCGGCTCAG 1024

    220 rProGlyAlaIleHisTyrCysAlaSerThrSerGlu.............. 232
        |      ::::::   ||||||:::   :::::::::
   1025 CTCCGTCAGCGTCACCTACTGCTCTGTCAGTCAGGACTTCCCAGGCAGCA 1074

    232                                          ........... 232

   1075 ACTTGAATTTGCTCACCAACAATTCTGGGACTCCCAAAGACCACGACTCC 1124

    233 ..................................................GluSe 234
                                                          :::||
   1125 CCTGAGAACGGTGCGGACAGCTTCGAGAGCTCAGACTCCCTCCTCCAGTC 1174

    234 rTrpThrAsp.....................................S 238
        ||||:::::::                                  :
   1175 CTGGAACAGCCAGTCGTCCTTGCTGGATGTGCAACGGGTTCCTTCCTTCG 1224

    238 erGluValAspSerSerCysSer.......................... 245
        :::::   ::::::::::||||||
   1225 AGAGCTTCGAAGATGACTGCAGCCAGTCTCTCTGCCTCAATAAGCCAACC 1274

    245 ..................................................... 245

   1275 ATGTCTTTCAAGGATTACATCCAAGAGAGGAGTGACCCAGTGGAGCAAGG 1324

    246                                        ........GlyGlnP 248
                                                       :::   |
   1325 CAAACCAGTTATACCTGCAGCTGTGCTGGCCGGCTTCACAGGAAGTGGAC 1374

    248 roIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSer 264
        ||||:::|||||||||||||||   |||   ||||||:::       |||
   1375 CTATTCAGCTGTGGCAGTTTCTCCTGGAG   CTGCTATCAGACAAATCC 1421

    265 TyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIl 281
                ||||||   |||   ...   ...          |||||||::
   1422 TGCCAGTCATTCATCAGCTGGACTGGA   GACGGATGGGAGTTTAAGCT 1468

    281 eGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgP 298
        :   |||       :::||||||||   ||||||   ||||||||||:::|
   1469 CGCCGACCCCGATCAGCTGGCCCGCCGGTGGGGAAAGAGGAAAAATAAGC 1518

    298 roAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLys 314
        ||   |||||||||:::|||||||||||||||:::::|||   ||||||
   1519 CCAACATGAACTACGAGAAGCTGAGCCGCGGCTTACGCTACTATTACGAC 1568

    315 LysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPh 331
        |||:::|||||:::|||   :::   :::::|||   ||||||:::||
   1569 AAGAACATCATCCACAAGACGTCG   GGCAAGCGCTACGTGTACCGCTT 1615

    331 eVal 332
        ||||
   1616 CGTG 1619

seq name: /SIDS8/gcgdata/geneseq/published/DNA2000/DAT/AAA39419

seq_documentation_block:
ID   AAA39419 standard; cDNA; 2269 BP.
XX
AC   AAA39419;
```

```
XX
DT   04 SEP 2000  (first entry)
XX
DE   Human ets2 coding sequence.
XX
KW   Human; ets2; chromosome 21q22.3; protooncogene; transcription factor;
KW   cell cycle control, cell proliferation, cyclin D1, cdc 2, liver cancer,
KW   ovarian cancer; breast cancer; lung cancer; bladder cancer;
KW   colo rectal cancer; rectal cancer; prostate cancer, cervical cancer, ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             292..1761
FT                   /*tag= a
FT                   /product= "Human ets2"
XX
PN   WO200060590 A2.
XX
PD   02 JUN 2000.
XX
PF   24 NOV 1999;   99WO US27805.
XX
PR   25 NOV 1998;   98US 0109850.
XX
PA   (MUSC ) MUSC FOUND RES DEV.
PA   (PAPA/) PAPAS T C.
XX
PI   Watson DK;
XX
DR   WPI; 2000 399904/34.
DR   P PSDB; AAY98040.
XX
PT   Novel modified ets2 protein, antisense molecules and ribozyme
PT   molecules, useful for detecting and treating cancer, especially
PT   prostate cancer
XX
PS   Example 1; Fig 6; 101pp; English.
XX
CC   The present sequence is the human ets2 coding sequence. The human ets2
CC   gene is localised to chromosome 21q22.3. The ets2 gene is a proto-oncogene
CC   and its gene product functions as a transcription factor. Ets2 protein is
CC   involved in cell cycle control and has been shown to a regulator of cdc-2
CC   and cyclin D1. Derequlation of  expression has been implicated in a
CC   number of cancers, especially prostate and cervical. Inhibition of ets2
CC   function in cancer cells leads to a reduction of the transformed
CC   properties of the cancer cells. Therefore, inhibition of ets2 function is
CC   a potential target for cancer therapy. An antisense ets2, a ribozyme
CC   molecule, or a triplex molecule, and compounds that decreases the
CC   expression, or antagonise the activity of ets2 may be used to treat
CC   cancer (preferably liver, ovarian, breast, lung, bladder, kidney,
CC   colo rectal, cervical or prostate cancer). In addition, a modified
CC   version of the ets2 protein may be produced, in which the transcription
CC   activation region is deleted and preferably replaced by a repressor
CC   element.
XX
SQ   Sequence 2269 BP; 550 A; 643 C; 584 G; 492 T; 0 other;


alignment scores:
            Quality:  299.00              Length:    552
              Ratio:  1.484                 Gaps:     18
 Percent Similarity:  39.130   Percent Identity: 22.464

alignment block:
 US-09-126-945B-2 x AAA89419  ..

 Align seq 1/1  to: AAA89419  from: 1  to: 2269

        8 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 49
          |||:::  |||      |||:::|||  ||||||     |||||||||
        6 PCCPCCCCPCCCCPCCACPCGGCCGTCCCTCCTTCCTCCPCCCPCCPCCC 55

       19 oProAspThrValSerArqThrGlyLeuGluLysAlaAlaAlaGlyAlaV 36
          ||||      :::  :::          |||:::||||||  |||:::
       56 TCCTCCPCCCGCTCCTGAA.........GAGCGCGCCGCGTGGGGGACCG 96

       36 alGlyLeuGluArqArqAspTrpSerProSerProProAlaThrProGlu 52
          ::  |||                          ||||||  |||  |||
       97 CCCGGTTA.....................CTTCCTCCAGAGACTGACGAG 125

       53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPhe......AspMetLeuTy 67
             |||:::::              |||              :::
      126 TGCGGTGTCGCTCCAGCTCAGAGCTCCCGGAGCCGCCCGGCCAGCGTCCG 175

       67 rProGluAspSerSer.TrpAlaAlaLysAlaPro..............  78
          |||      ||||| |||   |||   :::|||
      176 GCCTCCCTGATCGTCTCTGGCCGGCGCCCTCGCCCTCGCCCGCCGCGCAC 225

       79 ...GlyAlaSerSerArqGluGluProProGluGluProGluGlnCysPr 94
             :::|||:::::   |||:::||||||::::::|||:::      ||
      226 CCACCAGCCCCCGCCCCACCAGCCACCGTCCCCACCAACCGCCCCCCC 275

       94 oVal.....................................IleAspSerG 99
          |:::                                  :::|||:::
      276 TGCCCGCAGCGGCAGGATGAATGATTCGGAATCAAGAATATGGACCAGG 325

       99 lnAlaProAlaGly..................................... 104
          ||||||:::::
      326 TAGCCCCTGTGGCTAACAGTTACAGAGGGACACTCAAGCGCCAGCCAGCC 375

      104 ..................................................... 104

      376 TTTGACACCTTGATGGGTCCCTGTTTGCTGTTTTTCCTTCTCTAAATGA 425

      104 .......SerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisS 118
                :::|||:::   ||||||  ||||||        :::|||:
      426 AGAGCAAACACTGCAAGAAGTGCCAACAGGCTTGGATTCCATTCTTCATG 475

      118 or.......................LeuGluGlnValGlnSerMetVal 126
          ::                      |||         ::::::  |||
      476 ACTCCGCCAACTGTGAATTGCCTTTGTTAACCCCGTGCAGCAAGGCTGTG 525

      127 ValGlyGlaValLeuLysAspIleGluThrAlaCysLys........... 139
          :::::::::::::||||||        ::::::   |||
      526 ATGAGTCAAGCCTTAAAAGCTACCTTCAGTGGCTTCAAAAAGGAACAGCG 575

      140  LeuLeuAspIleThrAlaAspProMetAspTrpSerProSerAsnValG 156
           |||:::|||      :::|||     ||||||  :::::|||
      576 GCGCCTGGGCATTCCAAAGAACCCCTGGCTGTGGAGTGAGCAACAGGTAT 625

      156 lnLysTrpLeuLeuTrpThrGluHisGluTyrArqLeuProProMetGly 172
          :::||||||||||||:::  :::::::::  |||     ::::::
      626 GCCAGTGGCTTCTCTGGGCCACCAATGAGTTCAGTCTGGTGAACGTGAAT 675

      173 LysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGl 189
                  :::  |||:::  ||||||  ::::::::::||
      676 CTGCAGAGGTTCGGCATGAATGGCCAGATGCTGTGTAACCTTGGCAAGGA 725

      189 uGlnPheArqGlnArqSerPro...LeuGlyGlyAspValLeuHisAlaH 205
          |:::|||  :::  ::|||  :::  |||||::::|||     |
      726 ACGCTTTCTGGAGCTGGCACCTGACTTTGTGGGTGACATTCTCTGGGAAC 775

      205 isLeuAsp.............................................. 207
          |||||:::
      776 ATCTGGAGCAAATGATCAAAGAAAACCAAGAAAAGACAGAAGATCAATAT 825

      208 ..............HetTrpLysSerAlaAlaTrpMet............ 219
                       :::  :::        |||:::
      826 GAAGAAAATTCACACCTCACCTCCGTTCCTCATTGGATTAACAGCAATAC 875

      219 ..................................................... 219

      876 ATTAGGTTTTGGCACAGAGCAGGCGCCCTATGGAATGCAGACACAGAATT 925

      219 ..................................................... 219
```

```
      926 ACCCCAAAGGCGGCCTCCTGGACAGCATGTGTCCGGCCTCCACACCCAGC 975

      216 .....LysGluArgThrSe 220
                                  |||:::|||   ||
      976 GTACTCAGTCTGAGCAGGAGTTTCAGATGTCCCCAAGTCTCGGCTCAG 1025

      220 rProGlyAlaIleHisTyrCysAlaSerThrSerGlu..... 232
          |    ::::::   |||||||:::   ::::::::::
     1026 CTCCGTCAGCGTCACCTACTGCTCTGTCAGTCAGGACTTCCCAGGCAGCA 1075

      232 ..... 232
     1076 ACTTGAATTTGCTCACCAACAATTCTGGGACTCCCAAAGACCACGACTCC 1125

      234 .....GluSe 234
                                  :::||
     1126 CCTGAGAACCCTGCCCACACCTTCAGACCTCAGACTCCCTCCTCCAGTC 1175

      234 rTrpThrAsp.....S 248
          ||||::::::                :
     1176 CTGGAACAGCCAGTCGTCCTTGCTGGATGTGCAACGGGTTCCTTCCTTCG 1225

      238 erGluValAspSerSerCysSer..... 253
          :::::   :::::::::||||||
     1226 AGAGTTCGAAGATGACTGCAGCCAGTCTCTCTGCCTCAATAAGCCAACC 1275

      245 ..... 245
     1276 ATGTCTTTCAAGGATTACATCCAAGAGACGAGTGACCCAGTGGAGCAAGG 1325

      246 .....GlyGlnP 248
                                  :::   |
     1326 CAAACCAGTTATACCTGCAGCTGTGCTGGCCGGCTTCACAGGAAGTGGAC 1375

      248 roIleHisLeuTrpGlnPheLeuLysGluLeuLeuLysProHisSer 283
          ||||:::|||||||||||||||   |||   ||||||:::      |||
     1376 CTATTCAGCTGTGGCAGTTTCTCCTGGAG   CTGCTATCAGACAAATCC 1422

      265 TyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysII 281
                ||||||   |||   :::   :::          ||||||::
     1423 TGCCAGTCATTCATCAGCTGGACTGGA...GACGGATGGGAGTTTAAGCT 1469

      281 eGlnAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgP 298
          :   |||      :::||||||||||   ||||||   |||||||||:::|
     1470 CGCCGACCCCGATGAGGTGGCCGCCGGTGGGGAAAGAGGAAAAATAAGC 1519

      298 roAlaMetAsnTyrAspLysLeuSerArgSerIleArgLeuTyrTyrLys 314
          ||   |||||||:::|||||||||||||:::::|||   ||||||
     1520 CCAAGATGAACTACGAGAAGCTGAGCCGGGGCTTACGCTACTATTACGAC 1569

      315 LysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPh 331
          |||:::|||||:::|||   :::   :::::|||   ||||||:::||
     1570 AAGAACATCATCCACAAGACGTCG...GGGAAGCGCTACGTGTACCGCTT 1616

      331 eVal 332
          ||||
     1617 CGTG 1620

seq_name; /S1:e8/ipdata/_pto_seq_protseq_NA1999_DAT_AA18*.*.

seq_documentation_block:
ID   AAZ18359 standard; cDNA; 2428 BP.
XX
AC   AAZ18359;
XX
DT   19-OCT-1999  (first entry)
XX
DE   Human ASTHl1 alt1 cDNA sequence.
XX
KW   ASTH1; asthma; human; chromosome 11p; ASTH11; ASTH1J; genetic locus;
KW   therapeutic; immunogen; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             189..986
FT                   /*tag   a
FT                   /product= "ASTH11 alt1 protein"
XX
PN   WO9937809-A1.
XX
PD   29-JUL-1999.
XX
PF   21 JAN 1998,   98WO-US01260.
XX
PR   21-JAN-1998;   98WO-US01260.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Brooks Wilson AF,  Bohlen A,  Gordon E,  Carey AH;
PI   Galvin M,  Miller A,  North M;
XX
DR   WPI; 1999-479058/40.
DR   P-PSDB; AAY31417.
XX
PT   Mammalian asthma related genes, useful for diagnosis of a
PT   predisposition to development of asthma
XX
PS   Claim 15; Page 104-105; 195pp; English.
XX
CC   The invention identifies a genetic locus ASTH1, associated with asthma,
CC   mapped to human chromosome 11p. ASTH11 and ASTH1J are genes present
CC   within the locus, located close to each other on human chromosome 11p,
CC   and have similar patterns of expression, and common sequence motifs. The
CC   ASTH1 genes and fragments, encoded protein, genomic regulatory regions
CC   and anti-ASTH1 antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for the modulation of gene
CC   activity in vivo, for prophylactic and therapeutic purposes. The ASTH1
CC   protein is useful as an immunogen to raise specific antibodies, in drug
CC   screening for compositions that mimic or modulate ASTH1 activity or
CC   expression, including altered forms of ASTH1 protein, and as a
CC   therapeutic. The present sequence represents a human ASTH11 alt1
CC   cDNA sequence.
XX
SQ   Sequence 2428 BP; 705 A; 519 C; 519 G; 685 T; 0 other;


alignment scores:
            Quality:  299.00              Length:    211
              Ratio:  2.418                 Gaps:      4
 Percent Similarity:  61.137   Percent Identity:  43.175

alignment block:
 US-09-126-945B-2 x AAZ18359   ..

 Align seq 1/1  to: AAZ18359  from: 1  to: 2428

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
          :::||||||:||                :::::   |||     ||:::
      333 CAGACAGCCTGTGACTCATACTGGACATCACTCCACGCCTGAAIACTGGAC 382

      151 rTr SerAsnValGluLysTrpLeuGlyTrpThrGlnHisGlnTyrArgL 168
          :     :::|||   :::|||||   :::         |||||||:::|
      383 TAAGCGCCATCTGTGGGAGTGGCTCCACTTCTGCTGCGACCAGTACAACT 432

      168 euProProMetGlyLysAlaPhe.....GlnGluLeuAlaGlyLysGlu 182
          ||              :::|||         ::::::::::|||   :::
      433 TGGACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGGCCTGCAG 482

      183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199
          ||||||:::|||:::::::|||:::|||   :::   :::   |||   ||
      483 CTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAGCTGGCCTCTGCGG 532

      199 yAspValLeuHisAlaHisLeu..... 206
          |:::   |||:::      |||
```

```
      543 CGAGTACCTGTACTTCATCCTCCAGAACATCCGCACACAAGGTTACTCCT 592

      207 .......AspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSer 220
                  |||      :::|||  |||  :::|||:::  ::::::
      593 TTTTTAATGACGCTGAAGAAAGCAAGCCCACCATCAAAGACTATGCTGAT 642

      221 ProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAs 237
             :::          |||  :::|||||      :::      ||
      643 TCCAAC...........TGCTTGAAAACAAGTGGCATCAAAAGTCAAGA 670

      237 pSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnP 254
          |  :::     |||     |||  |||     ||||||||||:::|
      671 CTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGGAAT 720

      254 heLeuLysGluGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArg 270
          ||:::::::::|||||||||:::|||:::::      :::::::::
      721 TTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTGGAA 770

      271 TrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVa 287
          |||   :::::|||:::|||||||||:::::      |||     ::
      771 TGGGAACATAGGCAACAAGGAATTTTTCGGGTGGTTAAATCGGAACCCCT 820

      287 lAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspL 304
          :|||:::::::|||||   |||||:::::::      |||:::|||:::|
      821 GGCAAACATGTGGGGACAAAGCAAGAAAAATGACAGAATGACGTATGAAA 870

      304 ysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLys 320
          ||||||||||:::::|||  |||||||||  |||||:::::::::
      871 AGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGGAATTTTGGAGCGG 920

      321 ProAspIleSerGlnArgLeuValTyrGlnPhe 341
             |||      :::|||||||||||||:::|||
      921 GTTGAC......CGAAGGTTAGTGTACAAATTT 947

Seq name: /cds1/sequip/prodseq/prodseqB_NA.seq:AI_AAA*:1

seq documentation block:
ID   AAA80257 standard; cDNA; 2428 BP.
XX
AC   AAA80257;
XX
DT   22-NOV-2000 (first entry)
XX
DE   Human ASTHI1 cDNA, splice variant alt1.
XX
KW   ASTHI locus; ASTHI1; human; chromosome 11p; asthma;
KW   bronchial hyperreactivity; ets family; transcription factor;
KW   splice variant; genetic predisposition; polymorphism; antibody;
KW   drug screening; prophylaxis; therapy; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             189..986
FT                   /*tag= a
FT                   /product= "Human ASTHI1 isoform alt1"
XX
PN   US6087485-A.
XX
PD   11-JUL-2000.
XX
PF   21-JAN-1998;    98US-0009913.
XX
PR   21-JAN-1997;    97US-0035663.
PR   01-JUL-1997;    97WO-US011432.
XX
PA   (AXYS ) AXYS PHARM INC.
XX
PI   Galvin M;  Miller A;  North M;  Cardon L;  Buckler A;
PI   Brooks Wilson AR;  Carey AH;
XX
DR   WPI; 2000-505109/45.
DR   P-PSDB; AAB11969.
XX
PT   New nucleic acids other than naturally occurring chromosomes encoding
PT   ASTHI protein, for e.g. screening compositions that modulate expression
PT   or function of ASTHI proteins or as diagnostics for genetic
PT   predisposition to asthma -
XX
PS   Examples; Column 129-132; 131pp; English.
XX
CC   The invention relates to the ASTHI locus on the short arm of human
CC   chromosome (11p). This locus comprises the ASTHI1 and ASTHI2 genes,
CC   which are associated with a genetic predisposition to asthma and
CC   bronchial hyperreactivity. The ASTHI1 and ASTHI2 genes are oriented in
CC   opposite directions with the ASTHI locus, and have similar patterns of
CC   expression and common sequence motifs. They are both expressed in
CC   trachea, lung and several other tissues. ASTHI1 and ASTHI2 are novel
CC   members of the ets family of transcription factors, which have been
CC   implicated in the activation of a variety of genes including the ICAM
CC   gene and cytokine genes known to be important in the aetiology of
CC   asthma. Both ASTHI1 and ASTHI2 mRNAs are alternatively spliced.
CC   Alternative splicing of transcripts has no effect on the open reading
CC   frame of ASTHI2, as the exons involved are all 5' to the start codon in
CC   exon 6. In contrast, alternative splicing of ASTHI1 transcripts results
CC   in 3 different ASTHI1 isoforms. The invention also encompasses mouse
CC   asthi2 protein. The ASTHI nucleic acids are useful as diagnostics to
CC   identify a hereditary predisposition to asthma, as probes for identifying
CC   ASTHI related genes, for identifying expression of the gene in a
CC   biological specimen, and for generating genetically modified non human
CC   animals or site specific gene modifications in cell lines. The encoded
CC   ASTHI proteins are useful as immunogens to raise specific antibodies; in
CC   drug screening for compositions that mimic or modulate activity or
CC   expression of ASTHI and/or ASTHI2 (including altered forms of these
CC   proteins); and as a therapeutic. The ASTHI genes or fragments thereof,
CC   encoded proteins, ASTHI genomic regulatory regions, and anti-ASTHI1 and
CC   anti-ASTHI2 antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for modulation of gene activity
CC   in vivo for prophylactic and therapeutic purposes. The intact ASTHI1 or
CC   ASTHI2 proteins or active fragments thereof may be used to modulate or
CC   reduce bronchial hyperreactivity. Sequences AAA80257-A80259 represent
CC   three alternatively spliced human ASTHI1 cDNAs which encode three
CC   different isoforms of ASTHI1 (AAB11969-B11971).
XX
SQ   Sequence 2428 BP; 705 A; 519 C; 519 G; 685 T; 0 other;


alignment_scores:
            Quality:  299.00              Length:    211
              Ratio:  2.818                 Gaps:      4
 Percent Similarity:  61.137   Percent Identity: 33.175

alignment_block:
 US-09-126-945B-2 x AAA80257   ..

 Align seq 1/1  to: AAA80257  from: 1  to: 2428

      145 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
          :::|||||||||                 ::::::   |||      ||:::
      333 CAGACAGCCTGTGACTCATACTGGACATCAGTCCACCCTGAATACTGGAC 382

      151 rProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 168
          :      :::|||   :::|||||   :::         |||||||:::|
      383 TAAGCGCCATGTCTGGGAGTGGCTCCAGTTCTGCTGCGACCACTACAAGT 432

      168 euProProMetGlyLysAlaPhe......GlnGluLeuAlaGlyLysGlu 182
          ||          :::|||          ::::::::::|||    :::
      433 TGGACACCAATTGCATCTCCTTCTGAACTTCAACATCAGTGGCCTGCAG 482

      183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199
          |||||:::|||:::::::||:::|||   :::   :::   |||   ||
      483 CTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAGCCTGGCCTCTGCGG 532

      199 yAspValLeuHisAlaHisLeu..................... 206
          |:::   ||||::      |||
      533 CGAGTACCTGTACTTCATCCTCCAGAACATCCGCACACAAGGTTACTCCT 582
```

```
      207 .......AspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSer 220
                  |||      ::|||   |||   :::|||:::   ::::::
      583 TTTTAATGACGCTGAAGAAAGCAAGGCCACCATCAAAGACTATGCTGAT 632

      221 Pr GlyAlaIleHisTyr?,sAlaS:ThrSerGluGluSerTrpThrAs 267
             :::            |||   :::|||||||      :::      ||
      633 TCCAAC............TGCTTGAAAACAAGTGGCATCAAAAGTCAAGA 676

      237 pSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnP 254
          |   :::      |||      |||   |||      ||||||||||:::|
      671 CTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGGAAT 720

      254 heLeuLysGlnLeuLeuLeuLysProHisSerTyrGlyArgPheIleArg 270
          |:::::::::|||||||||:::|||:::::      :::::::::
      721 TTGTACGACACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTGGAA 770

      271 TrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGluVa 287
          |||   :::::|||:::|||||||||:::::      |||      ::
      771 TGGGAAGATAGGGAACAAGGAATTTTTCGGGTCGTTAAATCGGAAGCCCT 820

      287 lAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspL 304
          :|||:::::|||||   |||||:::::      |:|:::|||:::|
      821 GGCAAAGATGTGGGGACAAACGAAGAAAAATGACACAATGACGTATGAAA 870

      304 ysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLys 320
          |||||:|||||:::::::||   |||||||||   |||||:::::::::
      871 AGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGGAATTTTGGATCGG 920

      321 ProAspIleSerGlnArgLeuValTyrGlnPhe 331
            |||      :::|||||||||||||:::|||
      921 GTTGAC.....CGAAGGTTAGTGTACAAATTT 947

seq_name: [illegible]

seq_documentation_block:
ID   AAZ98756 standard; DNA; 1090 BP.
XX
AC   AAZ98756;
XX
DT   20-JUN-2000  (first entry)
XX
DE   Murine mammary specific transcription factor BrEts1 nucleotide sequence.
XX
KW   BrEts; transcription factor; mammary gland; milk protein synthesis; ds;
KW   lactation; dairy cattle; milk production; breast cancer; immunosuppressant;
KW   prevent; mouse.
XX
OS   Mus sp.
XX
PN   WO200012552 A1.
XX
PD   09-MAR-2000.
XX
PF   30-AUG-1999;   99WO-NZ00144.
XX
PR   28-AUG-1998;   98NZ-0331612.
XX
PA   (LHUI/) L'HUILLIER P.J.
PA   (LANG/) LANGLEY B.
PA   (BROP/) BROPHY B.
XX
PT   L'Huillier P.J.,  Langley B.,  Brophy B;
XX
DR   WPI; 2000-256590/22.
XX
PT   New polypeptide designated BrEts is a transcription factor useful for
PT   altering milk production or composition in mammals, and for immunising,
PT   treating or preventing breast cancer -
XX
PS   Disclosure; Fig 1; 32pp; English.
XX
CC   This sequence represents the murine mammary specific transcription
CC   factor BrEts1 nucleotide sequence. The transcription factor is isolated
CC   from mammary tissue and in mice BrEts expression is switched on at day
CC   4-6 of gestation and continues throughout pregnancy and lactation until
CC   day 2 of involution. The BrEts expression pattern correlates with the
CC   transcription of milk protein genes such as beta-casein, indicating that
CC   BrEts may be a switch for milk protein gene expression. The
CC   transcription factor may be used in a method for modulating the
CC   development of a mammary gland. BrEts may alternatively be used in a
CC   method for manipulating milk protein synthesis in cultured cells or in a
CC   transgenic mammal, to enhance the level of milk production, alter milk
CC   composition and/or the timing of milk protein synthesis. The invention is
CC   used to alter milk production levels or composition in mammals or
CC   cultured cells, to aid selection for milk production or animal
CC   performance in dairy cattle, and to diagnose, treat or prevent breast
CC   cancer. Note: This sequence is encoded by a murine BrEts cDNA sequence;
CC   it is unclear in the specification if the sequence is represented by
CC   AAZ98756 or AAZ98757.
XX
SQ   Sequence 1090 BP; 284 A; 292 C; 253 G; 260 T; 1 other;


alignment_scores:
            Quality:  296.00              Length:    207
              Ratio:  2.260                 Gaps:      5
 Percent Similarity:  63.285   Percent Identity:  33.333

alignment_block:
 US-09-126-945B-2 x AAZ98756  ..

 Align seq 1/1  to: AAZ98756  from: 1  to: 1090

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
          :::|||||||||               ::::::   |||      ||||::
      [illegible] 286

      151 rProSerAsnValGluLysTrpLeuLeuTrpIhrGluHisGluTyrArgL 168
          :      :::|||   :::|||||||   :::          ||||||:::|
      287 CAAGCGCCAIGTCTGGGAATGGCTCCAATTCTGCTGTGACCAGTACAAGC 336

      168 euProProMetGlyLysAlaPhe......GlnGluLeuAlaGlyLysGlu 182
          ||              :::|||          ||||||||||||   :::
      337 TTGATGCCAACTGCATCTCCTTCTGTCACTTCAACATCAGCGGCCTGCAG 486

      183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199
          |||||:::|||||||||||||:|||   :::   :::   :::   ||
      487 CTCTGTAGCATGAGTGAGGAGCAGTTCATTGAGCGAGCGGGCATCTGTGG 436

      199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMet. 215
          |:::   |||:::      |||:::      :::|||   ::::::
      437 GGAGTACCTGTACTTCATTCTCCAGAACATTCGCTCGCAAGGTTACTCCT 486

      216 ..........LysGluArgThrSerProGlyAlaIleHisTyr...... 226
                    |||:::|||:::   |||            |||
      487 TTTTCAATGATGCTGAAGACGACCAAGACTGGCATCAAAGACTATGCTGAT 536

      227 ......CysAlaSerThrSerGluGluSerTrpThrAspSerGluValAs 241
                ||||   :::|||||||      :::      :::::::::
      537 TCCAGTTGCTTGAAAACAAGT......GGCATCAAGAGTCAAGACTGTCA 580

      241 pSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluL 258
          |||      |||   |||      ||||||||||:::|||:::::::::|
      581 CAGCCGAACAAGCCTCCAAAGTTCTCATCTATGGGAATTTGTCAGAGACT 630

      258 euLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 274
          |||||||||:::||,:::::      :::::::::|||   :::::
      631 TGCTGCTGTCCCCTGAAGACGAACTGTGGCATCCTGGAATGGGAAGACAGG 680

      275 GluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTr 291
          |||:::|||||||||:::::      |||      :::|||:::::::||
      681 GAGCAGGGCATTTTCCGACTGGTTAAGTCAGAAGCCCTGGCAAAGATGTG 730

      291 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 308
          ||||   ||||||:::::      |||:::|||:::|||||||||||||:
```

```
      741 GGGACAAAGGAAGAAGAATGACAGGATGACGTACGAGAAGCTGAGCCGAG 790

      308 erIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSer 324
          :::::|||   ||||||||   ||||||:::::::   |||
      781 CCCTGACATACTACTATAAAACGGGAATTCTGGAGCGGGTTGAC..... 824

      325 GlnArgLeuValTyrGlnPhe 331
          :::||||||||||||:::|||
      825 CGGAGGTTAGTGTACAAATTT 845

Seq name: /SIEVE/projdata/geneseq/genseqp2/NA2000.DAT:AAZ98757

Seq documentation block:
ID   AAZ98757 standard; DNA; 2196 BP.
XX
AC   AAZ98757;
XX
DT   20-JUN-2000 (first entry)
XX
DE   Murine mammary specific transcription factor BrEts2 nucleotide sequence.
XX
KW   BrEts; transcription factor; mammary gland; milk protein synthesis; ds;
KW   lactation; dairy cattle; milk production; breast cancer; diagnose; treat;
KW   prevent; mouse.
XX
OS   Mus sp.
XX
PN   WO200012552-A1.
XX
PD   09-MAR-2000.
XX
PF   30-AUG-1999;  99WO-NZ00144.
XX
PR   28-AUG-1998;  98NZ-0331612.
XX
PA   (LHUI/) L'HUILLIER P.J.
PA   (LANG/) LANGLEY B.
PA   (BROP/) BROPHY B.
XX
PI   L'Huillier P.J., Langley B., Brophy B;
XX
DR   WPI; 2000-256590/22.
XX
PT   New polypeptide designated BrEts is a transcription factor useful for
PT   altering milk production or composition in mammals, and for diagnosing,
PT   treating or preventing breast cancer
XX
PS   Disclosure; Fig 1; 32pp; English.
XX
CC   This sequence represents the murine mammary specific transcription
CC   factor BrEts2 nucleotide sequence. The transcription factor is isolated
CC   from mammary tissue and in mice BrEts expression is switched on at day
CC   4-6 of gestation and continues throughout pregnancy and lactation until
CC   day 2 of involution. The BrEts expression pattern correlates with the
CC   transcription of milk protein genes such as beta-casein, indicating that
CC   BrEts may be a switch for milk protein gene expression. The
CC   transcription factor may be used in a method for modulating the
CC   development of a mammary gland. BrEts may alternatively be used in a
CC   method for manipulating milk protein synthesis in cultured cells or in a
CC   transgenic mammal, to enhance the level of milk production, alter milk
CC   composition and/or the [illegible] milk protein [illegible]. The [illegible] is
CC   used to alter milk production levels or composition in mammals or
CC   cultured cells, to aid selection for milk production or animal
CC   performance in dairy cattle, and to diagnose, treat or prevent breast
CC   cancer. Note: This sequence is encoded by a murine BrEts cDNA sequence,
CC   it is unclear in the specification if the sequence is represented by
CC   AAZ98756 or AAZ98757.
XX
SQ   Sequence 2196 BP; 597 A; 524 C; 476 G; 597 T; 2 other;


alignment scores:
            Quality:  296.00              Length:    207
                Ratio:  2.260                 Gaps:      5
 Percent Similarity:  68.285   Percent Identity:  44.333

alignment block:
 US-09-126-945B-2 x AAZ98757  ..

 Align seq 1/1  to: AAZ98757  from: 1  to: 2196

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
          :::||||||||||                ::::::   |||      ||:::
      199 CAGACAGCCTGTCATCCTACTGGACATCAGTGCACCTGAATACTGGAC 248

      151 rProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 168
          :      ::::|||   :::|||||||   :::          ||||||:::|
      249 CAAGCGCCATGTCTGGGAATGGCTCCAATTCTGCCTGTGACCAGTACAAGC 298

      168 cnProProMetGlyLysAlaPhe......GlnGluLeuAlaGlyLysGlu 182
          ||              :::|||          ::::::::::||   :::
      299 TTGATGCCAACTGCATCTCCTCTGTCACTTCAACATCAGCGGCCTGCAG 348

      183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199
          ||||||:::|||:::::|||:::|||   :::   :::   :::   ||
      349 CTCTGCAGCATGACGCAGGAGGAGTTCATTGAGGCAGCCGGCATCTGTGG 398

      199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMet. 215
          |:::   ||:::      |||:::      :::|||   ::::::
      399 GGAGTACCTGTACTTCATTCTCCAGAACATTCGCTCGCAAGGTTACTCCT 448

      216 ..........LysGluArgThrSerProGlyAlaIleHisPyr..... 226
                    |||:::|||:::   |||         |||
      449 TTTTCAATGATGCTGAAGAGACCAAGACTGGCATCAAAGACTATGCTGAT 498

      227 ......CysAlaSerThrSerGluGluSerTrpThrAspSerGluValAs 241
                |||   :::|||||||      :::      :::::::::
      499 TCCAGTTGCTTGAAAACAAGT......GGCATCAAGAGTCAAGACTGTCA 542

      241 pSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGlnL 258
           |||      |||   |||      ||||||||||:::|||:::::::::|
      543 CAGCCGAACAAGCCTCCAAAGTTCTCACCTGTGGGAATTTGTCAGAGACT 592

      259 euLeuLeuLysPr0HisSerTyrGlyArgProIleArgTrpLeuAsnLys 274
          ||||||||:::|||:::::::        :::::::::|||   ::::::
      593 TGCTGCTGTCCCCTGAAGAGAACTGTGGCATCCTGGAATGGGAAGACAGG 642

      275 GluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTr 291
          |||:::||||||||:::::::      |||      :::|||:::::::|
      643 CAGCAGGGCATTTTCCGACTGGTTAAGTCAGAAGCCCTGGCAAAGATGTG 692

      291 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 308
          ||||   ||||||::::::      |||:::|||:::|||||||||||:
      693 GGGACAAAGGAAGAAGAATGACAGGATGACGTACGAGAAGCTGAGCCGAG 742

      308 erIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSer 324
          :::::|||   ||||||||   ||||||:::::::   |||
      743 CCCTGACATACTACTATAAAACGGGAATTCTGGAGCGGGTTGAC..... 786

      325 GlnArgLeuValTyrGlnPhe 331
          :::||||||||||||:::|||
      787 CGGAGGTTAGTGTACAAATTT 807

 Query sequence: /SIEVE/projdata/geneseq/genseqp2/NA2000.DAT:AAZ98758

Seq documentation block:
ID   AAZ98758 standard; DNA; 983 BP.
XX
AC   AAZ98758;
XX
DT   20-JUN-2000 (first entry)
XX
DE   Bovine mammary specific transcription factor BrEts nucleotide sequence.
XX
KW   BrEts; transcription factor; mammary gland; milk protein synthesis; ds;
KW   lactation; dairy cattle; milk production; breast cancer; diagnose; treat;
```

```
KW   prevent.
XX
OS   Bos sp.
XX
PN   WO200012552-A1.
XX
PD   09-MAR-2000.
XX
PF   30-AUG-1999;   99WO-NZ00144.
XX
PR   28-AUG-1998;   98NZ-331612.
XX
PA   (LHUI/) L'HUILLIER P.J.
PA   (LANG/) LANGLEY B.
PA   (BROP/) BROPHY B.
XX
PI   L'Huillier P.J., Langley B., Brophy B.;
XX
DR   WPI; 2000-255990/22.
DR   P-PSDB; AAY29065.
XX
PT   New polypeptide designated BrEts is a transcription factor useful for
PT   altering milk production or composition in mammals, and for diagnosing,
PT   treating or preventing breast cancer
XX
PS   Disclosure; Fig 1; 32pp; English.
XX
CC   This sequence represents the bovine mammary specific transcription
CC   factor BrEts nucleotide sequence. The transcription factor is isolated
CC   from mammary tissue and in mice BrEts expression is switched on at day
CC   4-6 of gestation and continues throughout pregnancy and lactation until
CC   day 2 of involution. The BrEts expression pattern correlates with the
CC   transcription of milk protein genes such as beta-casein, indicating that
CC   BrEts may be a switch for milk protein gene expression. The
CC   transcription factor may be used in a method for modulating the
CC   development of a mammary gland. BrEts may alternatively be used in a
CC   method for manipulating milk protein synthesis in cultured cells or in a
CC   transgenic mammal, to enhance the level of milk production, alter milk
CC   composition and/or the timing of milk protein synthesis. The invention
CC   used to alter milk production levels or composition in mammals or
CC   cultured cells, to aid selection for milk production or animal
CC   performance in dairy cattle, and to diagnose, treat or prevent breast
CC   cancer. Note: This sequence is encoded by a murine BrEts cDNA sequence,
CC   it is unclear in the specification if the sequence is represented by
CC   AAZ98756 or AAZ98757.
XX
SQ   Sequence 984 BP; 257 A; 241 C; 250 G; 235 T; 0 other;


alignment_scores:
            Quality:  295.00              Length:     207
              Ratio:  2.323                 Gaps:       3
 Percent Similarity:  61.353   Percent Identity:  42.850

alignment_block:
 US-09-126-945B-2 x AAZ98758

 Align seq 1/1  to: AAZ98758  from: 1  to: 984

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
          :::||||||||||                ::::::   |||       |:|::
      251 CAGACAGCTTGCGACTCCTACTGGACATCCGTCCACCCCGAATACTGGAC 300

      151 rProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 169
          :      ::::||   :::||||||   :::          |||||:::|
      301 GAAGCGCCACGTCTGGGAATGGCTCCAGTTCTGCTCTCGACCACTACAAGC 350

      169 euProProMetGlyLysAlaPhe......GlnGluLeuAlaGlyLysGlu 182
          ||              :::|||         ::::::::|||   :::
      351 IGGACGCCAACTGCATCTCTTCTGCCATTTCAACATCAGTGGCCTGCAG 400

      183 LeuCysAlaMetSerGluGlnInPheArgGlnArgSerProLeuGlyS: 199
          |.||||::|||::::::|||:::|||   :::   :::   :::   ||

      401 CTGTGCGGCATGACACAGGAGCAGTTCATGGAGGCGCCGGGCGTCTGTGG 450

      199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrp.... 214
          |:::   |||:::      |||:::      :::|||   ::::::
      451 GGAGTATCTGTACTTTATCCTCCAGAGCATCCGCTCACAAGGTTACTCCT 500

      215 .....MetLysGluArgThrSerProGlyAlaIle 224
                |||          :::      |||
      501 TTTTAATGATCCTGATGAGACCAAGGCCACCCTCAAAGACTATGCTGAT 550

      225 HisTyrCysAlaSerThrSerGluGlnSerTrpThrAspSerGluValAs 241
          |||   :::||||||      :::      |||   :::
      551 TCCACTTGCTTGAAAACAAGCTGCATCAAAAGTCAACACTGTCACAGTCA 600

      241 pSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluL 258
          |||      |||   |||      |||||||||:::|||:::::::::|
      601 TAGTCGAACAAGCCTCCACAGTTCTCATCTATGGGAATTTGTCCGAGATC 650

      259 euLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 274
          ||||||||:::|||:::::          ::::::::|||   :::::
      651 IGCTTCTATCTCCTGAGGAAAACTGCGGCATTCTGGAATGGGAAGATAGG 700

      275 GluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTr 291
          |||:::|||||||||:::::      |||      :::|||:::::||
      701 GAACAAGGTATTTTTGGGGTGCTTAAATCACAAGCCCTGCCGAAGATGTG 750

      291 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 308
          ||||   ||||||:::::      |||:::|||:::|||||||||||:
      751 GGGACAAAGGAAGAAAAATGACAGAATGACGTACGAAAAGCTGAGCAGAG 800

      308 erIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSer 324
          :::::|||   |||||||||   ||||||:::::::   |||
      801 CTCTGAGGIACTACIATAAAACCGGAATTTTGGAACGGGTTGAC...... 844

      325 GlnArgLeuValTyrGlnPhe 331
          :::|||||||||||||:::|||
      845 CGAAGATTAGTGTACAAATTT 865

seq_number  SEQ 58 application number=pctnumber=pct/NZ1999 PAT AAX26551

seq_documentation_block:
ID   AAX26551 standard; DNA; 1447 BP.
XX
AC   AAX26551;
XX
DT   14-JUN-1999  (first entry)
XX
DE   DNA encoding chicken C-11 protein.
XX
KW   Chicken; C-11 protein; cell calcification inhibiting activity;
KW   cell calcification inhibiting agent; c-erg protein; arthritis deformans;
KW   ossification; spinal column ligament; ss.
XX
OS   Gallus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..1418
FT                   /*tag   a
XX
PN   JP11075871-A.
XX
PD   23-MAR-1999.
XX
PF   29-MAY-1998;   98JP-0166076.
XX
PR   20-JUN-1997;   97US-0050297.
PR   18-JUN-1997;   97US-0878177.
XX
PA   (CHUS ) CHUGAI PHARM CO LTD.
PA   (UYPE-) UNIV PENNSYLVANIA.
XX
DR   WPI; 1999-257708/22.
DR   P-PSDB; AAY01520.
```

```
XX
PF   An active protein for inhibiting cell calcification - useful for
PI   measuring the calcification of a cell, for diagnosing arthritis
PT   deformans or ossification of spinal column ligament
XX
PS   Disclosure; Page 7 8; 15pp; Japanese.
XX
CC   The present sequence encodes a chicken C 11 protein which has cell
CC   calcification inhibiting activity. The specification also describes
CC   a cell calcification inhibiting agent containing c erq protein
CC   (AAY01521). The proteins are used for measuring the calcification of a
CC   cell, for diagnosing arthritis deformans or ossification of spinal column
CC   ligament.
XX
SQ   Sequence 1447 BP; 440 A; 374 C; 317 G; 316 T; 0 other;


alignment scores:
            Quality:  294.50              Length:     846
              Ratio:  1.758                 Gaps:      14
 Percent Similarity:  48.555   Percent Identity: 27.457

alignment block:
 US-09-126-945B-2 x AAX26552   ..

 Align seq 1/1  to: AAX26552  from: 1  to: 1447

      67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 83
         |||   :::   ||::::   :::   ::::::::|||      :::
     183 TATGGCAAACATCAAAGATGAGCCCCGCGCTTCCCCAGCAGGACTGCTT 232

      83 qGluGluProPro..................................... 87
         ::::::||||||
     233 ATCACAGCCCCCGGCCAGAGTTACCATTAAGATGGAGTGTAACCCAAACC 282

      88 ..................GluGluProGluGlnCysProValIleAspSer 98
                           ::::::|||::::::|||   |||      :::
     283 AGGTTAATGGGTCAAGGAATTCACCTGATGACTGCAGCGTGGCAAAAGGA 332

      99 GlnAlaProAlaGlySerLeuAspLeuValPro.........GlyGlyIle 112
                 ::::::|||   |||   |||             |||:::
     333 GGGAAAATGGTTAGCAGTTCAGACAATGTTGGGATGAACTATGGAAGCTA 382

     112 uThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyG 129
               |||:::|||                  :::|||:::::::::|
     383 CATGGAAGAGAAGCATATT.....CCGCCTCCAAATATCACAACCAATG 426

     129 luValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAla 145
         ||                        :::   :::   :::   |||
     427 AA.........................GGAAGAGTTATTGTGCCAGCA 469

     146 AspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpTh 162
         ||||||      ||||||   ::::::|||::::::|||||   |||::
     470 GATCCTACGTTATGGAGCACAGACCATGTACGGCAGTGGCTGGAGTGGGC 499

     162 rGluHisGlnTyrArgLeuProProMetGly...LysAlaPheGlnGluL 178
         :     :::|||   ||||||   :::          ||||||:::::
     500 AGTGAAGGAGTATGGTCTTCCAGACGTGGACATCTTGTTGTTCCAGAACA 549

     178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGluArg 194
         ::   ||||||||||||||   |||:::::::::::::::::::::::
     550 TTGATGGCAAAGAGTTGTGTAAAATGACCAAAGATGACTTCCAGAGACTC 599

     195 SerPro...LeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLy 210
         :::|||      ::::::|||:::|||   :::||||||
     600 ACGCCGAGCTATAACGCAGATATCCTCCTGTCACACCTACAC........ 641

     210 sSerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIleHisTyrC 227
                 ::::::::::||||||          ||||||
     642 .........TACCTCAGAGAGAGA.........GGAGCCACTTTTATTT 672

     227 ysAlaSerThrSer.................................... 231
            :::|||||||

     673 TTCCAAATACATCAGTTTACCCAGAAGCAACGCAAAGAATAACAACAAGC 722

     232 ..........................GluGluSerTrp............ 235
                                    ::::::::::|||
     723 CCAGATTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGACGAGTCACAC 772

     235 ..................................................... 235

     773 CCATCCCACTCAGTCAAAAGCTACCCAACCATCATCTTCAACAGTGCCCA 822

     235 ..................................................... 235

     823 AAACAGAAGACCAGCGTCCTCAGTTAGATCCTTATCAGATTCTTGGACCG 872

     236 ThrAspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTr 252
         |||::::|||::::::   :::      |||||||||   |||:::|||||
     873 ACCAGCAGCCGTCTTGCAAATCCAGGGAGTGGGCAG...ATACAGCTATG 919

     252 pGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPhel 269
         |||||||||   |||   |||||:::      |||   ::::::   |
     920 GCAGTCCTACTGGAG...CTTCTGTCGGACAGCTCCAACTCCAACTGCA 966

     269 leArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAla 285
         ||   |||   :::      :::|||   |||||:::   |||
     967 TCACCTGG...GAGGCACAAATGGGGAGTTCAAGATGACAGACCCTGAT 1013

     286 GlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTy 302
         :::|||||||||   |||||   |||||:::::::|||   |||||||
    1014 GAAGTGGCTCGGCGTGGGAGAGAGGAAAAGCAAACTAACATGAACTA 1063

     302 rAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleA 319
         ||||||||||||||||:::::|||   ||||||   |||:::|||:::
    1064 TGACAAACTCAGCCGTGCACTTCGCTACTACTATGACAAAAATATTATGA 1113

     319 rgLysProAspIleSerGluArgLeuValTyrGlnPhe 351
           |||          :::::::|||   :::|||:::|||
    1114 CTAAAGTTCAT...GGTAAACGCTATGCCTACAAATTT 1148

seq name: /tmp/STN/tmp_seq_number_NA7999_AAX26552

seq documentation block:
ID   AAX26552 standard; DNA; 1528 BP.
XX
AC   AAX26552;
XX
DT   14-JUN-1999  (first entry)
XX
DE   DNA encoding chicken c erq protein.
XX
KW   Chicken; C 11 protein; cell calcification inhibiting activity;
KW   cell calcification inhibiting agent; c erq protein; arthritis deformans;
KW   ossification; spinal column ligament; ss.
XX
OS   Gallus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             68..1499
FT                   /*tag= a
XX
PN   JP11075871-A.
XX
PD   23-MAR-1999.
XX
PF   29-MAY-1998;   98JP-0166076.
XX
PR   20-JUN-1997;   97US-0050297.
PR   18-JUN-1997;   97US-0878177.
XX
PA   (CHUS-) CHUGAI PHARM CO LTD.
PA   (UYPE-) UNIV PENNSYLVANIA.
XX
DR   WPI; 1999-257708/22.
DR   P-PSDB; AAY01521.
```

```
XX
PT   An active protein for inhibiting cell calcification - useful for
PT   measuring the calcification of a cell, for diagnosing arthritis
PT   deformans or ossification of spinal column ligament
XX
PS   Disclosure; Page 8-9; 15pp; Japanese.
XX
CC   The present sequence encodes a chicken c-erg protein. The specification
CC   also describes a chicken C-11 protein (AAY01520) which has cell
CC   calcification inhibiting activity and a cell calcification inhibiting
CC   agent containing c-erg protein. The proteins are used for measuring the
CC   calcification of a cell, for diagnosing arthritis deformans or
CC   ossification of spinal column ligament.
XX
SQ   Sequence 1528 BP; 464 A; 395 C; 330 G; 339 T; 0 other;


alignment_scores:
            Quality:  292.00              Length:     464
              Ratio:  1.708                 Gaps:      14
 Percent Similarity:  46.978   Percent Identity:  26.923

alignment_block:
 US-09-126-945B-2 x AAY26552   ..

 Align seq 1/1  to: AAY26552  from: 1  to: 1528

      67 IyrProGluAspSerSerIrpAlaAlaLysAlaProGlyAlaSerSerAr  83
         |||   :::   |||:::   :::   :::::::|||      :::
     183 TATGGGCAAACATCAAAGATGAGCCCGCGCGTTCCCCAGCAGGACTGCTT 232

      83 qGluGluProPro..................................... 87
          :::::|||||
     233 AICACAGCCCCCGGCCAGAGTTACCATTAAGATGGAGTGTAACCCAAACCC 282

      88 ................GluGluProGluGlnCysProValIleAspSer  98
                         :::::|||:::::::|||   |||      :::
     283 AGGTTAATGGCTCAAGGAATTCACCTGATGACTGCAGCGTGGCAAAAGGA 332

      99 GlnAlaProAlaGlySerLeuAspLeuValPro.........GlyGlyLe 112
               :::::::|||   |||   |||           ||:::
     333 GGCAAAATGGTTAGCAGTTCAGACAATGTTGGGATGAACTATGGAAGCTA 382

     112 uThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyG 129
             |||:::|||                ::::|||:::::::::|
     383 CATGGAAGAGAAGCATATT......CCGCCTCCAAATATGACAACCAATG 426

     129 laValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAla 145
         ||                      :::   :::   :::   |||
     427 AA........................CGAAGAGTTAATTGTGCCAGCA 449

     146 AspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpTh 162
         |:|:||      ||||||   :::::|||:::::::|||||   |||::
     450 GATCCTACGTTATGGAGCACAGACCATGTACGGCAGTGGCTGGACTGGGC 499

     162 rGluHisGlnTyrArgLeuProProMetGly...LysAlaPheGluGluL 178
         :      ::::||   ||||||   :::          |||||||:::
     500 AGTGAAGGAGTATGGTCTTCCAGACGTGGACATCTTGTTGTTCCAGAACA 549

     178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGluArg 194
         ::   |||||||||||||   |||:::::::::::::::::::::::
     550 TTGATGGGAAAGAGTTGTGTAAAATGACCAAAGATGACTTCCAGAGACGC 599

     195 SerPro...LeuGlyGlyAspValLeuHisAlaHisLeu........... 206
         :::|||      :::::::|||:::|||   :::||||||
     600 ACGCCGAGCTATAACGCAGATATCCTCCTGTCACACCTACACTACCTCAG 649

     207 ..............................AspIleTrpLys....... 216
                                      ||:::  |||
     650 AGAGACTCCTCTTCCACATTTGACTTCAGATGATGTTGATAAGCCTTAC 699

     211 .....SerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIleHis 225
              |||         |||      |||:::::  ||||||
```

```
     700 AAAACTCTCCACGGTTAATGCATGCTAGAAACACAGGAGGAGCCACTTTT 749

     226 TyrCysAlaSerThrSer                                 231
              :::||||||
     750 ATTTTTCCAAATACATCAGTTTACCCAGAAGCAACGCAAAGAATAACAAC 799

     232 ..............................GluGluSerTrp....... 235
                                       :::::::::|||
     800 AAGGCCAGATTTACCTTATGAGCAAGCGAGCAGATCAGCGTGGACGAGTC 849

     235 ................................................... 235

     850 ACAGCCATCCCACTCAGTCAAAAGCTACCCAACCATCATCTTCAACAGTG 899

     235 ................................................... 235

     900 CCCAAAACAGAAGACCAGCGTCCTCAGTTAGATCCTTATCAGATTCTTGG 949

     236 ....ThrAspSerGluValAspSerSerCysSerGlyGlnProIleHisL 251
             ||:::|||:::::   :::      |||||||||   |||:::|
     950 ACCGACCAGCAGCCGTCTTGCAAATCCAGGGAGTGGGCAG...ATACAGC 996

     251 euTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArg 267
         ||||||||||||||   |||   |||||||:::      |||   ::::::
     997 TATGGCAGTTCCTACTGGAG...CTTCTGTCGGACAGCTCCAACTCCAAC 1043

     268 PheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSe 284
         |||   |||   :::      :::|||   |||||||:::   |||
    1044 TGCATCACCTGG...GAGGGCACAAATGGGGAGTTCAAGATGACAGACCC 1090

     284 rAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetA 301
         :::||||||||||   ||||||   ||||||:::::|||   ||||
    1091 TGATCAAGTGGCTCGGCGTTGGGGAGACAGGAAAAGCAAACCTAACATGA 1140

     301 snTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIle 317
         ||||||||||||||||||||:::::|||   ||||||   |||:::|||
    1141 ACTATGACAAACTCAGCCGTGCACTTCGCTACTACTATGACAAAAATATT 1190

     318 IleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
         :::   |||         :::::|||   :::|||:::|||
    1191 ATGACTAAAGTTCAT...GGTAAACGCTATGCCTACAAATTT 1229

seq_name: /SIS8/gcgdata/geneseqbnewseqs/NA2000_DAT_AA294397

seq_documentation_block:
ID   AA294397 standard; cDNA; 1528 BP.
XX
AC   AA294397;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Mouse ELF5b cDNA.
XX
KW   ELF5; ELF5b; mouse; ETS protein; transcription factor;
KW   cell proliferation; cell differentiation; therapy; diagnosis;
KW   cancer, tumour; autoimmune disease; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             117..878
FT                   /*tag= a
FT   polyA_signal    1392..1397
FT                   /*tag= b
FT   polyA_signal    1493..1499
FT                   /*tag= c
XX
PN   WO200012695-A1.
XX
PD   09-MAR-2000.
XX
PF   26 AUG 1999;   99WO-AU00691.
XX
```

```
PR   27-AUG-1998;   98AU-0005512.
PR   30-SEP-1998;   98AU-0006252.
XX
PA   (MONU ) UNIV MONASH.
XX
PI   Kola I,  Zhou J;
XX
DR   WPI; 2000-270820/23.
DR   P-PSDB; AAY79290.
XX
PT   ELF5 protein, useful for diagnosis, antibody generation and as
PT   screening tool for agents capable of modulating transcriptional events
PT   during cellular functioning such as in tumorigenesis, comprises an ETS
PT   domain
XX
PS   Claim 11; Page 93-95; 114pp; English.
XX
CC   The present sequence of that of cDNA encoding novel mouse ELF5 (see
CC   AAY79290), defined as belonging to the ETS family of transcription
CC   factors due to its expression of an ETS domain which recognises and
CC   binds the purine-rich GGA(A/T) core motifs.  Murine ELF5 cDNA was
CC   initially isolated from an adult mouse lung cDNA library.  The
CC   full length coding sequence was then obtained by RT-PCR and RACE
CC   using rat placental RNA.  2 Predominant transcripts were identified
CC   in placental tissue, ELF5a and ELF5b, of approximately 2.5 and 1.5
CC   kb, respectively, which differed only in the 3' untranslated region.
CC   The present sequence is that of cDNA for the ELF5b transcript.  The
CC   identification of human and murine ELF5 permits the classification
CC   and rational design of a range of products for use in therapy,
CC   diagnosis and antibody generation involving, e.g., regulation of
CC   cellular functional activity such as cellular differentiation or
CC   cellular proliferation.  These therapeutic molecules may also act
CC   as either agonists or antagonists of ELF5 function and will be
CC   useful in cancer and autoimmune disease therapy.
XX
SQ   Sequence 1528 BP; 386 A; 486 C; 357 G; 299 T; 0 other;


alignment scores:
            Quality:  288.00              Length:     207
              Ratio:  2.215                 Gaps:       5
 Percent Similarity:  62.802   Percent Identity:  42.850

alignment block:
 US-09-126-945B-2 x AAZ94397    ..

 Align seq 1/1  to  AAZ94397  from  1  to  1528

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
          :::||||||||||                 ::::::   |||      |||::
      241 CAGACAGCCTGTGATTCCTACTGGACATCAGTGCCACCCTGAATACTGGAC 290

      151 rProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGluTyrArgL 168
          :      :::|||   :::||||||   :::         |||||||:::|
      291 CAAGCAGCCATGATCTGGGAATGGCTCCAATCTGCTGTGACCAGTACAAGC 340

      168 euProProMetGlyLysAlaPhe.....GlnGluLeuAlaGlyLysGlu 182
          ||          ::::||          ::::::::::|||   :::
      341 TTGATGCCAACTGCATCTCCTTCTGTCACTTCAACATCAGCGGCTGCAG 390

      183 LeuCysAlaMetSerGluGluGlnPheArgGluArgSerProLeuGlyGl 199
          ||||||::|||::::::|||::|||   :::   :::   :::   ||
      391 CTCTGCAGCATGAGCAGGAGGAGTTCATTGAGGCAGCCGGCATCTGTGGC 440

      199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMet. 215
          |:::   |||:::      |||:::      ::::||   ::::::
      441 GGAGTACCTGTACTTCATTCTCCAGAACATTCGCTGCCAAGGTTACTCCT 490

      216 ...........LysGluArgThrSerProGlyAlaIleHisTyr..... 229
                    |||::::|||::   |||          |||
      491 TTTTCAATGATGCTGAAGACACCAACACTGCCATCAAACACTATGCTGAT 540

      227 .....CysAlaSerThrSerGluGluSerTrpThrAspSerGluValAs 241
```

```
                |||   :::||||||      :::      :::::::::
      541 TCCAGTTGCTTGAAAACAAGT......GGCATCAAGAGTCAAGACTGTCA 574

      241 pSerSerCysSerGlyGlnProIleHisLeuTrpGluPheLeuLysGluL 258
           |||      |||   |||      |||||||||:::|||:::::::::|
      575 CAGCCGAACAAGCCTCCAAAGTTCTCACCTGTGGGAATTTGTCAGAGACT 624

      258 euLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 274
          ||||||||::::|||||:::      :::::::::|||   ::::::
      625 TGCTGCTGTCCCCTGAAGAGAACTGTGGCATCCTGGAATGGGAACACAGG 674

      275 GluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTr 291
          |||::::|||||||||:::::      |||      :::|||:::::||
      675 GAGCAGGGCATTTCCGAGTGGTTAAGTCAGAAGCCTGGCAAAGATGTG 724

      291 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 308
          ||||   ||||||:::::      |||::::|||::|||||||||||:
      725 GGGACAAAGGAAGAAGAATCACAGGATGACGTACGACAAGCTGAGCCGAG 774

      309 erIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSer 324
          :::::||   ||||||||      |||:::::::::   |||
      775 CCCTGAGATACTACTATAAAACGAGAATTCTGGAGCGGGTTGAC...... 818

      325 GlnArgLeuValTyrGluPhe 341
          :::|||||||||||||:::|||
      819 CGGAGGTTAGTGTACAAATTT 839

Db  Score  /site/mpi/data/seq/seqdata/NA2000.DAT:AAZ94396

Seq_documentation block:
ID   AAZ94396 standard; cDNA; 2224 BP.
XX
AC   AAZ94396;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Mouse ELF5a cDNA.
XX
KW   ELF5; ELF5a; mouse; ETS protein; transcription factor;
KW   cell proliferation; cell differentiation; therapy; diagnosis;
KW   cancer, tumour, autoimmune disease; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             117..878
FT                   /*tag   a
FT   polyA_signal    1392..1397
FT                   /*tag   b
FT   polyA_signal    1393..1399
FT                   /*tag   c
FT   polyA_signal    2181..2186
FT                   /*tag   d
FT   misc_feature    1721..1725
FT                   /*tag   e
FT                   /note  "A/T-rich tract"
FT   misc_feature    1878..1882
FT                   /*tag   f
FT                   /note  "A/T-rich tract"
FT   misc_feature    2072..2076
FT                   /*tag   g
FT                   /note  "A/T-rich tract"
FT   misc_feature    2091..2095
FT                   /*tag   h
FT                   /note  "A/T-rich tract"
XX
PN   WO200012695-A1.
XX
PD   09-MAR-2000.
XX
PF   26-AUG-1999;   99WO-AU00691.
XX
PR   27-AUG-1998;   98AU-0005512.
```

```
PP   30-SEP-1998;   98AU-000625Z.
XX
PA   (MONU ) UNIV MONASH.
XX
PI   Kola I,  Zhou J;
XX
DR   WPI; 2000-270820/23
DR   P-PSDB; AAY79290.
XX
PT   ELF5 protein, useful for diagnosis, antibody generation and as
PT   screening tool for agents capable of modulating transcriptional events
PT   during cellular functioning such as in tumorigenesis, comprises an ETS
PT   domain -
XX
PS   Claim 11; Page 91-93; 114pp; English.
XX
CC   The present sequence of that of cDNA encoding novel mouse ELF5 (see
CC   AAY79290), defined as belonging to the ETS family of transcription
CC   factors due to its expression of an ETS domain which recognises and
CC   binds the purine rich GGA(A/T) core motifs.  Murine ELF5 cDNA was
CC   initially isolated from an adult mouse lung cDNA library.  The
CC   full-length coding sequence was then obtained by RT-PCR and RACE
CC   using rat placental RNA.  2 Predominant transcripts were identified
CC   in placental tissue, ELF5a and ELF5b, of approximately 2.5 and 1.5
CC   kb, respectively, which differ only in the 3' untranslated region.
CC   The present sequence is that of cDNA for the ELF5a transcript.  The
CC   identification of human and murine ELF5s permits the identification
CC   and rational design of a range of products for use in therapy,
CC   diagnosis and antibody generation involving, e.g., regulation of
CC   cellular functional activity such as cellular differentiation or
CC   cellular proliferation.  These therapeutic molecules may also act
CC   as either agonists or antagonists of ELF5 function and will be
CC   useful in cancer and autoimmune disease therapy.
XX
SQ   Sequence 2224 BP; 603 A; 529 C; 481 G; 611 T; 0 other;


alignment_scores:
            Quality:  288.00              Length:    207
              Ratio:  2.215                 Gaps:      5
 Percent Similarity:  62.802   Percent Identity:  52.850

alignment_block:
 US-09-126-945B-2 x AAZ94396   ..

 Align seq 1/1  to: AAZ94396  from: 1  to: 2224

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
          :::|||||||||                ::::::   |||      ||||:
      231 CAGACAGCCTGTGATTCCTACTGGACATCAGTGCACCCTGAATACTGGAC 280

      151 rProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgI 169
          :      :::|||   :::|||||||   :::          |||||||:::|
      281 CAAGCGCCATGTCTGGCAATGGCTCCAATTCTGCTGTGACCAGTACAAGC 330

      168 euProProMetGlyLysAlaPhe      GlnGluLeuAlaGlyLysGlu 182
          ||           :::|||          ::::::::::|||   :::
      331 TTGATGCCAACTGCATCTCTTCTGTCACTTCAACATCAGCGGCCTGCAG 380

      183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199
          ||||||:::|||:::::::|||:::|||   :::   :::   :::   ||
      381 CTCTGCAGCATGAGCGAGGAGGAGTTCATTGAGGCAGCGGGCATCTGTGG 430

      199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMet. 215
          |:::   |||:::      |||:::      :::|||   ::::::
      431 GGAGTACCTGTACTTCATCTCCAGAACATTCGCTCGCAAGGTTACTCCT 480

      216            LysGluArgThrSerProGlyAlaIleHisTyr      229
                        |||:::|||:::   |||        |||
      481 TTTTCAAIGAIGCTGAAGAGACCAAGACTGGCATCAAAGACTATGCTGAT 530

      227 .....CysAlaSerThrSerGluGluSerTrpThrAspSerGluValAs 241
               |||   ::::||||||      :::      .........

      531 TCCAGTTGCTTGAAAACAAGT      GGCATCAAGAGTCAAGACTGTCA 574

      241 pSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluL 258
          |||      |||   |||      |||||||||:::|||:::::::::|
      575 CAGCCGAACAAGCCTCCAAAGTTCTCACCTGTGGCAATTTGTCAGAGACT 624

      259 euLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 274
          ||||||||:::|||:::::          :::::::::|||   ::::::
      625 TGCTGCTGAAGCCACGAAGACAACTGTGGCATCCTGGAATGGAAGACAGG 674

      275 GluLysGlyIlePheLysIleGlnAspSerAlaGlnValAlaArgLeuTr 291
          |||:::|||||||||:::::      |||      :::|||:::::::||
      675 GAGCAGGGCATTTTCCGAGTGGTTAAGTCAGAAGCCCTGGCAAAGATGTG 724

      291 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 308
          ||||   ||||||:::::      |||:::|||:::|||||||||||||:
      725 GGCACAAAGGAAGAAGAATCACAGGATGACGTACGAGAAGCTCAGCCGAG 774

      308 erIleArgGlnTyrGlyLysLysGlyIleIleArgLysProAspIleSer 324
          :::::|||   |||||||||      |||:::::::::   |||
      775 CCCTGAGATACTACATATAAAACGAGAATTCTGGAGCGGGTTGAC..... 818

      325 GlnArgLeuValTyrGlnPhe 331
          :::|||||||||||||:::|||
      819 CGGAGGTTAGTGTACAAATTT 839

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ50644

seq_documentation_block:
ID   AAQ50644 standard; cDNA; 2938 BP.
XX
AC   AAQ50644;
XX
DT   26-MAY-1994  (first entry)
XX
DE   Human Hum-Fli-1 gene clone BM025.
XX
KW   chromosomal translocation; chimeric; chimaeric; Ewing sarcoma;
KW   Ews gene; malignant melanoma; hum-fli-1;
KW   primitive peripheral neuroectodermal tumour; human chromosome 11;
KW   human chromosome 22; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             143  1501
FT                   /*tag=  a
FT                   /product= HUM-FLI-1
FT   polyA_signal    2908..2913
FT                   /*tag=  b
XX
PN   WO9323549-A.
XX
PD   25-NOV-1993.
XX
PF   19-MAY-1993;   93WO-FR00494
XX
PR   20-MAY-1992;   92FR-0006123.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Aurias A,  Delattre O,  Desmaze C,  Melot T,  Peter M;
PI   Plougastel B,  Thomas G,  Zucman J;
XX
DR   WPI; 1993-386580/48.
DR   P-PSDB; AAK44556.
XX
PT   New nucleic acid of EWS gene and its hybrid(s) - contg. gene
PT   sequence involved in chromosomal trans-location, also derived
PT   mRNA, probes, fusion proteins etc., for diagnosis and treatment
PT   of Ewing sarcoma and melanoma
XX
PS   Disclosure; Fig 7; 123pp; French.
```

```
XX
CC   The probe IIRR1 was used to screen a human marrow cDNA library
CC   (Clontech cat.# HL1058). The clone BM025 was identified and
CC   sequenced. It represents the entire coding region together with
CC   5' and 3' UTRs of the Hum Ets 1 gene.
XX
SQ   Sequence 2938 BP; 847 A; 692 C; 675 G; 724 T; 0 other;


Alignment Scores:
            Quality:  278.50              Length:     482
              Ratio:  1.522                 Gaps:      18
 Percent Similarity:  47.906   Percent Identity:  28.010

Alignment block:
 US-09-126-945B-2 x AAQ50644    ..

 Align seq 1/1  to: AAQ50644  from: 1  to: 2938

       34 GlyAlaValGlyLeuGluArg.....................Ar 41
          |||:::||||||   ||||||                      ||
      163 GGCTCTGTCGGTGGTGAGCGACGACCAGTCCCTCTTTGACTCAGCGTACG 212

       41 qAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerAlaP 58
          |:::   ||||||||||||               ||||||
      213 GACGCGCAGCCCATCTCCC.............CAAGGC.......... 252

       58 heTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAla 74
                          |||   |||
      253 ..........CGACATGACTGCCTC..................... 282

       75 AlaLysAlaProGlyAlaSer...SerArgGluGluProProGluGluPr 90
                     |||   |||      |||      |||   ::::::||
      283 ...........GGGGAGTCCTGACTACGGGCAGCCCCACAAGATCAACC 290

       90 qGluGlnCysProVal...IleAspSerGlnAlaProAlaGlySer.... 104
          |            :::   :::|||:::::|||      |||:::
      291 CCCTCCCACCACAGCAGGAGTGGATCAATCAGCCAGTGAGGGTCAACGTC 340

      105 .........................LeuAspLeuValProGlyGlyLeu 112
                                  ::::::   |||   |||||||||
      341 AAGCGGGAGTATGACCACATGAATGGATCCAGGGAGTCTCCGCTGGACTG 390

      113 .....................................ThrLeuGluGlnHi 117
                                             ::::::   ||
      391 CAGCGTTAGCAAATGCAGCAAGCTGGTGGGCGGAGGCGAGTCCAACCCCA 440

      117 sSerLeuGluGlnValGlnSerMetValVal.GlyGluValLeuLysAsp 133
          |:::|||:::|||:   :::          |||              :::
      441 TCAACTACAACAGCTATATGGACGAGAAGAATGGCCCCCCTCCTCCCAAC 490

      134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIr 150
          :::   |||      :::   :::   :::   |||||||||      ||
      491 ATGACCACCAACGAGAGGAGAGTCATCGTCCCCGCAGACCCCACACTGTG 540

      150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrA 167
          |:::   :::::||:::::::|||||   |||:::      :::|||
      541 GACACAGGAGCATGTGAGCCAATGGCTGGATGGGGCCATAAAGGACTATA 590

      167 rqLeuProProMetGlyLysAla...PheGlnGluLeuAlaGlyLysGlu 182
            |||      :::      :::   ||||||:::::   ||||||||
      591 GCTTGATGGAGATGGACACATCCTTTTTCCAGAACATGGATGGCAAGGAA 640

      183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeu...Gl 198
          ||||||   |||:::::::|||:::|||   :::   :::   |||   ::
      641 CTGTGTAAAATGAACAAGGAGGACTTCCTCCGCCGCCACCACCCTCTACAA 690

      198 yGlyAspValLeuHisAlaHisLeu                       206
          :   :::||||||   :::|||||
      691 CACGGAAGTGCTGTTGTCACACCTCAGTTACCTCAGGGAAAGTTCACTGA 740

      206 ................................................. 206
```

```
      741 TGGCCTATAATACAACCTCCCACACCGACCAATCCTCACGATTGAGTGTC 790

      207 ..................AspIleTrpLysSerAlaAlaTrpMetLysGl 217
                             |||      :::   :::||||||   :::::
      791 AAAGAAGACCCTTCTTATGACTCAGTCAGAAGAGGAGCATGGGGCAATAA 840

      217 sArgThrSerPro.....................GlyAlaIleHisTyrS 227
          :   :::|||                       ||||::   :::
      841 CATGAATTCTGGCCTCAACAAAAGTCCTCCCCTTGGAGGGGCACAAACGA 890

      227 ysAlaSerThrSerGluGla.............................. 233
            :::::::::::::|||:::
      891 TCAGTAAGAATACAGAGCAACGGCCCCAGCCAGATCCGTATCAGATCCTG 940

      234 SerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHi 250
          :::   ||:::::||::::::   :::      |||||||||   ||||:
      941 GGCCCGACCAGCAGTCGCCTAGCCAACCCTGGAAGCGGGCAG...ATCCA 987

      250 sLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyA 267
          :|||||||||||||||   |||   ||||||:::      :::   :::
      988 GCTGTGGCAATTCCTCCTGGAG...CTGCTCTCCGACAGCGCCAACGCCA 1034

      267 rqPheIleArqTrpLeuAsnLysGluLysGlyIlePheLysIleGluAsp 283
                |||   |||   :::      ::::|||   ||||||:::   |||
     1035 GCTGTATCACCTGG...GAGGGGACCAACGGGGAGTTCAAAATGACGGAC 1081

      284 SerAlaGlnValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMe 300
          :::|||||||||   ||||||   ||||||::::::|||   ||
     1082 CCCGATGAGGTGGCCAGGCGCTGGGGCGAGCGGAAAAGCAAGCCCAACAT 1131

      300 tAsnTyrAspLysLeuSerArqSerIleArqGlnTyrTyrLysLysGlyI 317
          ||||||||||||||||||||:::::|||   ||||||   ||:::|
     1132 GAATTACGACAAGCTGAGCCGGGCCCTCCGTTATTACTATGATAAAAACA 1181

      317 IeIleArqLysProAspIleSerGlnArqLeuValTyrGlnPhe 341
          ||::::   |||          :::::|||   :::|||:::||
     1182 TTATGACCAAAGTGCAC...GGCAAAAGATATGCTTACAAATTT 1222
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:AAQ37480

```
seq_documentation_block:
ID   AAQ37480 standard; DNA; 1985 BP.
XX
AC   AAQ37480;
XX
DT   30-JUN-1993  (first entry)
XX
DE   Sequence encoding purine-rich repeat (GA repeat) binding protien
DE   (GABP) subunit alpha.
XX
KW   GA binding protein; cis regulatory element;
KW   VP16 mediated induction; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             454..1818
FT                   /*tag   a
XX
PN   WO9304166-A.
XX
PD   04-MAR-1993.
XX
PF   17-AUG-1992;   92WO-US06748.
XX
PR   16-AUG-1991;   91US-0746042.
XX
PA   (CARN ) CARNEGIE INST WASHINGTON.
XX
PI   LaMarco KL,  Mc KnIght SL,  Thompson CC;
XX
DR   WPI; 1993-094998/11.
DR   P-PSDB; AAR43365.
```

```
XX
PT   DNA encoding GA binding protein sub-unit - allows investigation
PT   of sub-unit sequence motif functions, for control of rapid cell
PT   division e.g. in cancer
XX
PS   Claim 6; Fig 2A; 68pp; English.
XX
CC   A cis-regulatory element required for virion associated protein VP16
CC   mediated induction of herpes simplex virus 1 (HSV1) immediate early
CC   (IE) genes consists of three imperfect repeats of the purine-rich
CC   hexanucleotide 5' CGGAAR 3'. A protein complex capable of avid
CC   interaction with the purine-rich repeats (GA repeats) has been
CC   identified in soluble preparations of rat liver nuclei. This GA
CC   binding protein (GABP) consists of two separable subunits.
CC   Applicants have isolated cDNA clones encoding both subunits of GABP
CC   and have revealed that one (GABP alpha) is related to the Ets
CC   transforming protein, while the other (GABP beta) contains a
CC   series of 33-amino acid repeats related in sequence to a variety
CC   of proteins
XX
SQ   Sequence 1985 BP; 565 A; 441 C; 529 G; 450 T; 0 other;


alignment_scores:
            Quality:  277.00              Length:    289
              Ratio:  1.753                 Gaps:      9
 Percent Similarity:  54.861   Percent Identity: 28.819

alignment_block:
 US-09-126-945B-2 x AAQ87480  ..

 Align seq 1/1  to: AAQ87480  from: 1  to: 1985

      78 ProGlyAlaSerSerArgGluGlnProProGluGluProGluGlnCysPr 94
         |||  |||          |||          :::      |||||||||
     829 CCAGATCCCCACCAGCCCAACCAGAACCGCATCTCTTCAACAAGCTCA 878

      94 oValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyG 111
          ||||||                      :::||||||   :::
     879 AGTGATA.....................ACTCTTGACCGCACCAAGCAGA 907

     111 lyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValVal 127
          ||||||||||||||||||   |||||||||          ||||||
     908 TTACGACCATTTCAGACGACACCTCGCAGCAAGTGACGAGATGGGCTGCT 957

     128 GlyGlnValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleTh 144
          ::|||||||      :::      :::   |||:::|||
     958 .....GCACTGCAAGGCTACAGAAAAGAGCAGGAGCGCCTTGGCATCCC 1001

     144 rAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuT 161
          |||||||||   ||||||   :::::::|||      |||:::::::|
    1002 CTATGATCCTATACGCTGGTCCACGGACCAACTCCTGCATTGGGTGGTTT 1051

     161 rpIhrGluHisGlnTyrArqLeuProProMetGlyLysAlaPheGlnGlu 177
          ||:::      ::::::   :::      :::      :::      :::
    1052 GGGTAATGAAGGAGTTCAGCATGACTGATATAGACCTCACCACACTCAAC 1101

     178 LeuAlaGlyLysGluLeuCysAlaMetSerGlyGlnPheArgGlnAr 194
          ::::::|||:::|||||||||:::::::::::::|||:::|||   |||||
    1102 ATTTCGGGAAGAGAATTATGTAGTCTCAACCAAGAAGATTTTTTTCAGCG 1151

     194 qSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysS 211
          |   |||      |||:::::::|||   :::||||||:::::   ::::
    1152 GGTCCCT   CGGGGAGAAATTCTTTGGAGTCATCTGCAGCTTCTTCCAA 1198

     211 erAlaAlaTrpMetLysGluArq..................... 218
          ::   :::      :::::::::
    1199 AATATGTTTTGGCCAGCCAAGAGCAACAGATGAATGAGATAGTTACCATT 1248

     219 .....................................ThrSerProGl 232
                                             :::::|||
    1249 GACCAGCCTGTGCAGATTATTCCAGCCTCAGTGCCTCCCGCTACACGAC 1298

     222 yAlaIleHisTyrCysAlaSerThrSer.................GlnGlnS 234
          :::|||           |||::::::                 :::::::
    1299 TACGATTAAAGTTATAAACAGCAGTGCAAAAGCAGCTAAAGTGCAACGGT 1348

     234 erTrpThrAspSerGluValAspSerSerCysSerGlyGln..... 247
          ||      |||      |||   |||      ||:::
    1349 CCCCAAGGATTTCAGGAGAAGACAGAAGTTCACCGGGGAACAGAACAGGA 1398

     248 .........ProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuly 261
                  ||:::|||||||||||:|||   |||||||||
    1399 AACAATGGTCAGATCCAACTATGGCAGTTTTGCTAGAACTTCTTACTGA 1448

     261 sProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGly1 278
          :::             |||   |||::::::   |||   |||
    1449 CAAGGATGCTCGACAC...TGTATTTCTTGGGTTGGTGATGAA...GGTG 1492

     278 lePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArg 294
          ||||||:::::::::          |||||||::   ||||||   |||
    1493 AATTTAAGCTAAATCAGCCTGAGTTGGTTGCGCAAAAATGGGGACAACGT 1542

     295 LysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGl 311
          |||||||:::|||:::|||||||||||:::|||||||||||||:::::|||
    1543 AAGAACAAGCCTACCATGAACTATGACAAACTTAGCCGTGCATTACGGTA 1592

     311 nTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuV 328
          ||||||          :::|||   |||   :::   ::::::|||:::|
    1593 TTATTATCATGCGCACATGATTTCTAAAGTTCAA...GGCAAGAGATTTG 1639

     328 alTyrGlnPheVal 332
         |||||:::||||||
    1640 TGTACAAATTTGTT 1654

seq_name "gi|5174758|gb|U89673.1|HSU89673 Homo sapiens ASTHLJ mRNA" AAZ18358

seq_documentation_block:
ID   AAZ18358 standard; cDNA; 5667 BP.
XX
AC   AAZ18358;
XX
DT   19-OCT-1999  (first entry)
XX
DE   Human ASTHlJ alt3 cDNA sequence.
XX
KW   ASTH1; asthma; human; chromosome 11p; ASTH11; ASTH1J; genetic locus;
KW   therapeutic; immunogen; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/qualifiers
FT   CDS             390..1292
FT                   /*tag= a
FT                   /product= "ASTH1J protein"
XX
PN   WO9937809-A1.
XX
PD   29-JUL-1999.
XX
PF   21 JAN 1998.   98WO-US01260.
XX
PR   21-JAN-1998;   98WO-US01260.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Brooks-Wilson AR,  Buckler A,  Cardon L,  Carey AH;
PI   Galvin M,  Miller A,  North M;
XX
DR   WPI; 1999-479058/40.
DR   P-PSDB; AAY31416.
XX
PT   Mammalian asthma related genes, useful for diagnosis of a
PT   predisposition to development of asthma
XX
PS   Claim 15; Page 101-103; 195pp; English.
```

```
XX
CC   The invention identifies a genetic locus ASTH1, associated with asthma,
CC   mapped to human chromosome 11p. ASTH1I and ASTH1J are genes present
CC   within the locus, located close to each other on human chromosome 11p,
CC   and have similar patterns of expression, and common sequence motifs. The
CC   ASTH1 genes and fragments, encoded protein, genomic regulatory regions
CC   and anti ASTH1 antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for the modulation of gene
CC   activity in vivo for prophylactic and therapeutic purposes. The ASTH1
CC   protein is useful as an immunogen to raise specific antibodies, in drug
CC   screening for compositions that mimic or modulate ASTH1 activity or
CC   expression, including altered forms of ASTH1 protein, and as a
CC   therapeutic. The present sequence represents a human ASTH1J alt3
CC   cDNA sequence.
XX
SQ   Sequence 5667 BP; 1662 A; 1102 C; 1226 G; 1677 T; 0 other;


alignment scores:
            Quality:  266.00              Length:    428
              Ratio:  1.298                 Gaps:     18
 Percent Similarity:  47.897   Percent Identity: 25.467

alignment block:
 US-09-126-945B-2 x AAZ18858   ..

 Align seq 1/1  to: AAZ18858  from: 1  to: 5667

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
          :::|||:::   :::         :::   ||||||   :::|||:::||
       57 TTGGGAGGCATTGGCTCCACTGCAGCCTGGGTGTCTAGAGACCACATTCT 106

       17 u...LeuProProAspThrValSerArqThr.GlyLeuGluLysAlaAla 32
          |   ||||||            :::|||||| |||            :::
      107 CACCCTGCCTTTGTTACTGGCAAACCGAACGCGGCGCTGTGGCTTTCAGC 156

       33 AlaGlyAlaValGlyLeuGlu.........................Ar 40
             |||      |||||||::                          ||
      157 TTGGGTAAGCCGGGTCTGCGGCGGGGATTGCCATCTGAAGACAGAGGCAG 206

       40 qArqAspTrpSerProSerProProAlaThrPro.............. 51
          ||||      :::|||   |||   |||   |||
      207 GAGGGCAGCCACACCTTGCCCAGCTGCACACCCAGTAACAAGTTTCCTCA 256

       52 ..............................GluGlnGlyLeuSerAla 57
                                       ::::::||||||||||
      257 GTCCGGGTATCTGCCACAGGCTGGGCTGGTCATCAAAGGGCCTCAGTCAT 306

       58 PheTyrLeuSerTyrPheAspMetLeu.............TyrProGluAs 70
          :::   ::::::   |||      |||              |||   |||
      307 ATTTTAATAGAGCTCTTCAAGTATCTGGCTTTGTGATAATATCAGGAA.. 354

       70 pSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArqGluGluProP 87
                        :::   |||   |||::|||:::
      355 ...................TCAGTTGGTTTCTCTGACAGACACTG..... 400

       87 roGluGluProGluGlnCysProValIleAspSerGlnAlaProAlaGly 103
                           |||::|||                 ::::||
      401 ...................CCCATTATCATGATTCTGGAAGGAGGTGGT 410

      104 SerLeuAspLeuValProGlyGlyLeuThrLeu.................. 114
          :::::|||   ||||||::      |||
      411 GTAATGAATCTCAACCCCGGCAACAACCTCCTTCACCAGCCGCCAGCCTG 460

      115 ....GluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluV 130
              ::::::|||      :::|||:::||          |||
      461 GACAGACAGCTACTCCACGTGCAATGTTTCCAGTGGGTTTTTTGGAGGCC 510

      130 alLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAsp 146
                 :::|||:::
      511 AGTGGCATGAAATTCAT.................................. 527

      147 ProMetAspTrpSerProSerAsnValGluLysTrpLeuLeuTrpThrGl 163
          |||:::   |||:::      :::|||   :::||||||
      528 CCTCAGTACTGGACCAAGTACCAGTGTGGGAGTGGCTCCAGCACCTCCT 577

      163 uHisGlnTyrArqLeuProProMetGlyLysAlaPheGlnGlu......L 178
             :::|||                        |||||||||      :
      578 GGACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTTCGACA 627

      178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArqGlnArq 194
          ::   |||:::::::||||||:::||||||   :::::|||   :::
      628 TCAACGGCGAGCACCTCTGCAGCATGAGTTTGCAGGAGTTCACCGGGCG 677

      195 SerProLeuGlyGlyAspValLeuHisAlaHisLeu.............. 206
          :::       :::|||:::::::|||:::::::::::|||
      678 GCAGGGACGGCGGGGCAGCTCCTCTACAGCAACTTGCAGCATCTGAAGTG 727

      207 ....................AspIleIrpLysSer.........AlaAlaI 214
                              |||:::::::::|||         :::
      728 GAACGGCCAGTGCAGTACTGACCTGTTCCAGTCCACACACAATGTCATTG 777

      214 rpMetLysGluArqThrSerProGlyAlaIle................... 224
             |||:::|||:::|||:::   :::
      778 TCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTCGGAAAGACGAG 827

      225 HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluValAs 241
          :::|||     :::|||:::      |||   :::|||      :::||
      828 AACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG...TTGGA 874

      241 pSerSer......CysSerGlyGln........................... 247
          ||||:::      |||   :::|||
      875 CAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAGTCACC 924

      248 ..ProIle.............................................. 249
            |||:::
      925 TTCCTGTTGCAGAGTCACCTGATATCAAAAAGGAGCAAGACCCCCCTGCC 974

      250 .....................................HisLeuTrpGlnPh 254
                                              ||||||||||::||
      975 AAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGGAATT 1024

      254 eLeuLysGlnLeuLeuLeuLysProHisSerTyrGlyArqPheIleArqT 271
          |:::::::::::::|||||||::|||   :::          :::|||:::|
     1025 CATCGCGCAGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATAAAAT 1074

      271 rpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVal 287
          ||   :::::::::::::|||:::|||::::::        |||      |||
     1075 GGCAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGCAGTG 1124

      288 AlaArqLeuIrpGlyIleArqLysAsnArqProAlaMetAsnTyrAspLy 304
          |||:::||||||||   :::||||||:::   :::|||:::||||:::||
     1125 GCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTATGAAAA 1174

      304 sLeuSerArqSerIleArqGlnTyrTyrLysLysGlyIleIleArqLysP 321
          ||||||||||:::::|||   ||||||||||:::   |||:::::::::
     1175 GCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATACTGGAGCGTG 1224

      321 roAspIleSerGlnArqLeuValTyrGlnPhe 341
            |||   :::::|||||||||||||:::|||
     1225 TGGAT...GGACGAAGACTGGTATATAAATTT 1254

seq_name: /Seq_B/ProData/Seq/genseq_pt/NA2000_1AT.AAA80256

seq_documentation block:
ID   AAA80256 standard; cDNA; 5667 BP.
XX
AC   AAA80256;
XX
DT   22-NOV-2000  (first entry)
XX
DE   Human ASTH1J cDNA, splice variant alt3.
XX
KW   ASTH1 locus; ASTH1J; human; chromosome 11p; asthma;
KW   bronchial hyperreactivity; ets family; transcription factor;
KW   splice variant; genetic predisposition; polymorphism; antibody;
```

```
KW   drug screening; prophylaxis; therapy; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             390..1292
FT                   /*tag= a
FT                   /product= "Human ASTHL1"
XX
PN   US6087485 A.
XX
PD   11-JUL-2000.
XX
PF   21-JAN-1998;    98US-0009913.
XX
PR   21-JAN-1997;    97US-0035663.
PR   01-JUL-1997;    97US-0051442.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  North M,  Cardon L,  Buckler A;
PI   Brooks Wilson AR,  Carey AH;
XX
DR   WPI; 2000-505109/45.
DR   P-PSDB; AAB11968.
XX
PT   New nucleic acids other than naturally occurring chromosomes encoding
PT   ASTHl protein, for e.g. screening compositions that modulate expression
PT   or function of ASTHl proteins or as diagnostics for genetic
PT   predisposition to asthma -
XX
PS   Claim 5; Column 121-126; 131pp; English.
XX
CC   The invention relates to the ASTHl locus on the short arm of human
CC   chromosome (11p). This locus comprises the ASTHL1 and ASTHL2 genes,
CC   which are associated with a genetic predisposition to asthma and
CC   bronchial hyperreactivity. The ASTHL1 and ASTHL2 genes are oriented in
CC   opposite directions with the ASTHl locus, and have similar patterns of
CC   expression and common sequence motifs. They are both expressed in
CC   trachea, lung and several other tissues. ASTHL1 and ASTHL2 are novel
CC   members of the ets family of transcription factors, which have been
CC   implicated in the activation of a variety of genes including the TCRa
CC   gene and cytokine genes known to be important in the aetiology of
CC   asthma. Both ASTHL1 and ASTHL2 mRNAs are alternatively spliced.
CC   Alternative splicing of transcripts has no effect on the open reading
CC   frame of ASTHL1, as the exons involved are all 5' to the start codon in
CC   exon b. In contrast, alternative splicing of ASTHL1 transcripts results
CC   in 3 different ASTHL1 isoforms. The invention also encompasses mouse
CC   asthl2 protein. The ASTHl nucleic acids are useful as diagnostics to
CC   identify a hereditary predisposition to asthma, as probes for identifying
CC   ASTHl related genes, for identifying expression of the gene in a
CC   biological specimen, and for generating genetically modified non-human
CC   animals or site-specific gene modifications in cell lines. The encoded
CC   ASTHl proteins are useful as immunogens to raise specific antibodies; in
CC   drug screening for compositions that mimic or modulate activity or
CC   expression of ASTHL1 and/or ASTHL2 (including altered forms of these
CC   proteins); and as a therapeutic. The ASTHl genes or fragments thereof,
CC   encoded proteins, ASTHl genomic regulatory regions, and anti-ASTHL1 and
CC   anti-ASTHL2 antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for modulation of gene activity
CC   in vivo for prophylactic and therapeutic purposes. The intact ASTHL1 or
CC   ASTHL2 proteins or active fragments thereof may be used to modulate or
CC   reduce bronchial hyperreactivity. Sequences AAA80254-A80256 represent
CC   three alternatively spliced human ASTHL2 cDNAs.
XX
SQ   Sequence 5667 BP; 1662 A; 1102 C; 1226 G; 1677 T; 0 other;
```

```
alignment_scores:
            Quality:  266.00              Length:    428
              Ratio:  1.298                 Gaps:     18
 Percent Similarity:  47.897   Percent Identity: 25.467
```

```
alignment_block:
 US-09-126-945B-2 x AAA80256    ..

 Align seq 1/1  to: AAA80256  from: 1  to: 5667

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
          :::|||:::   :::       :::  ||||||  :::|||:::|| 
       57 TTGGGAGGCATTGGCTCCACTGCAGCCTGGGTGTCTAGAGACCACATTCT 106

       17 u...LeuProProAspThrValSerArgIhr.GlyLeuGluLysAlaAla 32
          |   |||||           :::|||||| |||          :::
      107 CACCCTGCCTTTGTTACTGGGAAACCGAACGCCGCCTGTGGCTTTCAGC 156

       33 AlaGlyAlaValGlyLeuGlu.........................Ar 40
             |||      |||||||::                        ||
      157 TTGGGTAAGCCGGGTCTGCGGCGGGGATTGCCATCTGAAGACAGAGGCAG 206

       40 qArgAspTrpSerProSerProProAlaThrPro.............. 51
          ||||      :::|||  |||  |||  |||
      207 GAGGGCAGCCACACCTTGCCCAGCTGCACACCCAGTAACAAGTTCCTCA 256

       52 ..............................GluGlnGlyLeuSerAla 57
                                       ::::::||||||||||
      257 GTGCGGGTATCTGCCACAGGCTGGGCTGGTCATCAAAGGGCCTCAGTCAT 306

       58 PheTyrLeuSerTyrPheAspMetLeu.............TyrProGluAs 70
          :::   ::::::   |||       |||           |||   |||
      307 ATTTAATAGAGCTCTTCAAGTATCTGGCTTTGTGATAATATCAGGAA.. 354

       70 pSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluProP 87
                          :::  |||  |||::::|||::
      355 ..................TCAGTTGGTTTCTCTGACAGACACTG..... 380

       87 roGluGluProGluGlnCysProValIleAspSerGlnAlaProAlaGly 103
                             |||::::||           ::::||
      381 ....................CCCATTATCATGATTCTGGAAGGAGGTGGT 410

      104 SerLeuAspLeuValProGlyGlyLeuThrLeu.................. 114
          ::::::|||   |||||:::      |||
      411 GTAATGAATCTCAACCCCGGCAACAACCTCCTTCACCAGCCGCCAGCCTG 460

      115     GluGluHisSerLeuGlpGlnValGlnSerMetValValGlyGluV 130
              ::::::::|||      :::|||::|||         |||
      461 GACAGACAGCTACTCCACGTGTAATGTTCCACTGGGTTTTTGGAGGCC 510

      130 alLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAsp 146
                ::::|||:::
      511 AGTGGCATGAAATTCAT................................ 527

      147 ProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuIrpIhrGl 163
          ||||::   |||:::       :::|||   :::|||||
      528 CCTCAGTACTGCACCAAGTACCACGTGTGGGAGTGGCTCCAGCACCTCCT 577

      163 uHisGlnTyrArqLeuProProMetGlyLysAlaPheGlnGlu......L 178
                  ::::||                 |||||||||      :
      578 GCACACCAACCACGTCCATCCAATTGTATCCCTTTCCAACACTTCCACA 627

      178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArqGlnArq 194
          ::   |||::::::|||||||:::|||||   ::::::|||   :::
      628 TCAACGGCGAGCACCTCTGCAGCATGAGTTTGTAGGAGTTCACCCGGGCG 677

      195 SerProLeuGlyGlyAspValLeuHisAlaHisLeu.............. 206
          :::     :::|||::::::|||:::::::::|||
      678 GCAGGCACGGCGGGGCAGCTCCTCTACAGCAACTTGCAGCATCTGAAGTG 727

      207 ...................AspIleIrpLysSer.........AlaAlal 214
                             |||:::::::::|||         :::
      728 GAACGGCCAGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGTCATTG 777

      214 rpMetLysGlnArqIhrSerProGlyAlaIle................... 224
                |||:::|||:::|||:::   :::
      778 TCAACACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAGACGAG 827
```

```
     225 HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluValAs 241
         :::|||      :::|||:::      |||   :::|||      :::||
     828 AACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG...TTGGA 874

     241 pSerSer......CysSerGlyGln..................... 247
         ||||:::      |||   :::|||
     875 CAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAGTCACC 924

     248 ..Profile..................................... 249
           |||:::
     925 TTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCCCTGCC 974

     250 .....................................HisLeuTrpGlnPh 264
                                             ||||||||||:::||
     975 AAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGGAATT 1024

     254 eLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgT 271
         |:::::::::::::|||||||:|||   :::         :::||||:::|
    1025 CATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATAAAAI 1074

     271 rpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVal 287
         ||   :::::::::::::|||:::|||::::::      |||      |||
    1075 GGGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGCAGTG 1124

     288 AlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLy 304
         ||||:::|||||||||   :::|||||:::   :::|||:::|||:::||
    1125 GCTCAGCTATGGGCTAAAAACGAAGAACAACAGCAGCATGACCTATGAAAA 1174

     304 sLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysP 321
         ||||||||||:::::|||   |||||||||||:::   |||:::::::::
    1175 GCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATACTGGAGCGTG 1224

     321 roAspIleSerGlnArgLeuValTyrGlnPhe 341
           |||   ::::::|||||||||||||:|||
    1225 TGGAT...GGACGAAGACTGGTATATAAATTT 1253

seq name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DA1:AAF92427

seq documentation block:
ID   AAF92427 standard; cDNA; 873 BP.
XX
AC   AAF92427;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Bovine mammary tissue derived cDNA #40.
XX
KW   Bovine; mammary gland; cancer; tumour; angiogenesis; ss.
XX
OS   Bos taurus.
XX
PN   WO200114554-A1.
XX
PD   01-MAR-2001.
XX
PF   24-AUG-2000; 2000WO-NZ00166.
XX
PR   24-AUG-1999;  99US-0150430.
XX
PA   (GENE ) GENESIS RES & DEV CORP LTD.
PA   (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.
XX
PI   Havukkala I.J,  Gleen M,  Grigor MR,  Molenaar AJ;
XX
DR   WPI; 2001-226619/23.
XX
PT   New polypeptides and polynucleotides encoding the polypeptides, which
PT   are expressed in bovine mammary gland tissue, useful for stimulating
PT   mammary gland growth or function, or inducing differentiation of milk
PT   producing cells.
XX
PS   Claim 1; Page 52; 97pp; English.
XX
CC   The present invention relates to proteins derived from bovine
CC   mammary gland cells. The invention is useful for stimulating
CC   bovine mammary gland cell growth and function, inhibiting the
CC   growth of various mammary gland cancer cells, inhibiting
CC   angiogenesis and vascularization of tumours, or modulating
CC   the growth of blood vessels in a mammal.
XX
SQ   Sequence 873 BP; 237 A; 213 C; 213 G; 210 T; 0 other;


alignment_scores:
            Quality:  265.50              Length:    201
              Ratio:  2.289                 Gaps:      4
 Percent Similarity:  57.711   Percent Identity:  40.846

alignment_block:
 US-09-126-945B-2 x AAF92427   ..

 Align_seq 1/1  to: AAF92427  from: 1  to: 873

     135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
         :::|||||||||      :::::  |||     ||||:
     200 CAGACAGCTGGCGACTCCTACTGGACATCCGTCCACCCCGAATACTGGAC 249

     151 rProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 168
         :      :::|||  :::||||||  :::       |||||||:|
     250 GAAGCGCCACGTCTGGCAATGGCTCCAGTTCTGCTGTGACCAGTACAAGC 299

     168 euProProMetGlyLysAlaPhe......GluGluLeuAlaGlyLysGl 182
         ||               :::|||         ::::::::::|||   :::
     300 TGGACGCCAACTGCATCTCTTTCTGCCATTTCAACATCAGTGGCCTGCAG 349

     183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199
         |||||:::|||:::::::::|||   :::|||             ||
     350 CTGTGCGGCATGACACAGGAGGAGTTCATGGAGCGGCCGGCGTCTGTGGG 399

     199 yAspVal......LeuHisAlaHisLeuAspIleTrpLysSerAlaAlaP 214
         |:::::      |||   :::::   :::      |||   :::   :
     400 GAGTATCTGTACTTTATCCTCCAGAGCATCCGCTCACAAGGTTACTCCTT 449

     214 rpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThr 230
         :::::           |||
     450 TTTTAATGATCCTGATGAGACCAAGGCCA..................... 478

     231 SerGluGluSerTrpThrAspSerGluValAspSerSerCysSerGlyGl 247
                                                 |||   ||
     479 .....................................AGCCTCCA 486

     247 nProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisS 264
         |      ||||||||:::|||:::::::::|||||||||:::||:::::
     487 GAGTTCTCATCTATGGGAATTTGTGCGAGATCTGCTTCTATCTCCTGAGG 536

     264 rTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLys 280
         ::         :::::::::|||      :::|||:::||||||||||:::
     537 AAAACTGCGGCATTCTGGAATGGGAAGCTAGGGAACAAGGTATTTTTCGG 586

     281 IleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnAr 297
         :::      |||      :::|||:::::|||||   ||||||:::::
     587 GTGGTTAAATCAGAAGCCCTGCCGAAGATGTGGGGACAAAGGAAGAAAAA 636

     297 gProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyrTyrL 314
         :      |||:::|||:::|||||||||||||:::::|||   |||||||
     637 TGACAGAATGACGTACGAAAAGCTGAGCAGAGCTCTGAGGTACTACTATA 686

     314 ysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGln 330
         ||   ||||||:::::::::   |||      :::|||||||||||:::
     687 AAACCGGAATTTTGGAACGGGTTGAC.....CGAAGATTAGTGTACAAA 730

     331 Phe 331
         |||
     731 TTT 733

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV15530
```

```
seq_documentation_block:
ID   AAV15530 standard; cDNA; 3240 BP.
XX
AC   AAV15530;
XX
DT   22-MAY-1998  (first entry)
XX
DE   Human lymphoid-specific transcription factor NERF-2 cDNA.
XX
KW   ets-related; human; lymphoid specific transcription factor; NERF-2;
KW   screening assay; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             207..1952
FT                   /*tag= a
FT                   /product= NERF-2
XX
PN   US5721113-A.
XX
PD   24-FEB-1998.
XX
PF   03-JAN-1995;   95US-0368281.
XX
PR   03-JAN-1995;   95US-0368281.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Kunsch CA,  Libermann TA,  Oettgen JP;
XX
DR   WPI; 1998-168405/15.
DR   P-PSDB; AAW47238.
XX
PT   NERF-1 and NERF-2 nucleic acids  encode ets-related human
PT   lymphoid-specific transcription factors
XX
PS   Claim 2; Columns 23-28; 17pp; English.
XX
CC   The present sequence encodes the ets-related lymphoid-specific
CC   transcription factor NERF-2, which can be used in screening assays
CC   for drugs that modulate NERF activity, and to treat patients having
CC   need of NERF-2. ets-related factors are involved in, e.g. cancer
CC   development, retrovirus replication, T-cell cytokine production,
CC   cell cycle effects, growth regulation and cell differentiation.
XX
SQ   Sequence 3240 BP; 1070 A; 592 C; 662 G; 916 T; 0 other;


alignment_scores:
            Quality:  263.00              Length:     415
              Ratio:  1.584                 Gaps:      11
 Percent Similarity:  48.116   Percent Identity: 24.058

alignment_block:
 US-09-126-945B-2 x AAV15530   ..

 Align seq 1/1  to: AAV15530  from: 1  to: 3240

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 56
         ||:::   ::::::   |||:::             :::::::||:::::
     192 CCTCAGGGAATAACAATGACATCAGCAGTGGTTGACAGTGGAGGTACTAT 241

      56 l.................GlyLeuGluArgArgAspTrpSer....... 44
         :                  |||:::|||::::::::::   :||
     242 TTTGGAGCTTTCCAGCAATGGAGTAGAAAATCAAGAGGAAAGTGATAAGG 291

      45 ............................ProSerProProAlaThrPro 51
                                     |||   |||   |||
     292 TTTCTGAATATCCAGCAGTGATTGTGGAGCCAGTTCCAAGTGCCACATTA 341

      52 GluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyr.. 67

         |||||||||   :::|||      |||::::||   |||      |||
     342 GAGCAGGGTATGCAGCCCAGGTTCTGGGTTATGATGATGAGACTTATTT 391

      68 .ProGluAspSerSerTrpAlaAlaCysAlaProGlyAlaSerSerArgG 84
          ::|||   :::                                     |
     392 GATGCAAGATGTGGCA.................................G 408

      84 IuGluProProGluGluProGluGlnCysProValIleAspSer..... 98
         |||||        |||   ||:::::   ::::::::::::
     409 AAGAACAACAAGTTGAGACCGAGAATGCGGAAACAGTGGAAGCATCAGTT 458

      99 .....................................GlnAlaProAlaGl 103
                                             ::|||      ::
     459 CACAGCAGTAATGCACACTGTACAGATAAGACAATTGAAGCTGCTGAAGC 508

     104 ySerLeuAspLeuValProGlyGlyLeuThrLeuGluGlnHisSerLeuC 120
         :   |||   :::                 ::::::::::|||   |
     509 CCTGCTTCATATGGAATCTCCTACCTGCTTGAGGGATTCAACAAGTCCTG 558

     120 IuGlnValGluSerMetValValGlyGluValLeuLysAspIleGluThr 136
         ||   :::::::::   :::      :::||:::              ||||||
     559 AATTCATCCATGCTGCTATGAGGCCAGATGTCATT......ACAGAAACT 602

     137 AlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSe 153
         :::   :::::   :::      ::::::||||||||   ||||||
     603 GTAGTGGAGGTGTCAACTGAAGAGTCTGAACCCATGGATACCTCTCCT.. 650

     153 rAsnValGluLysTrpLeuLeuTrpThrGluHisGluTyrArgLeuProP 170
                                                  ::::||
     651 .............................................ATTCCAA 657

     170 roMetGlyLysAlaPheGluGluLeuAlaGlyLysGluLeuCysAlaMet 186
                 :::   :::   :::      ||:::          :::
     658 CATCACCAGATAGCCATGAACCAATGAAAAAGAAAAAA.........GTT 698

     187 SerGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuHi 203
         :::::::::          |||:::||||||:::::::::
     699 GGCCGTAAACCAAAGACCCAGCAATCACCAATTTCCAATGGG........ 740

     203 sAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrS 220
                           :::                ::::|||::::::
     741 ...................TCTCCTGAGTTAGGTATAAAGAAGAAACCAA 771

     220 erProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThr 236
         |||
     772 GAGAAGGA........................................... 779

     237 AspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGl 253
                                    ::::|:::      :::|||||||::
     780 ........................AAAGGAAACACAACCTATTTGTGGGA 805

     253 nPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleA 270
         :|||||   :::||||||        :::::::      ||||:::|||:
     806 GTTTCTTTAGATCTACTTCAAGATAAAAATACTTGTCCCAGGTATATTA 855

     270 rgTrpLeuAsnLysGluGlyIleIleLysIleGluAspSerAlaGln 286
         ::|||   ::::::|||||||||||||||||||:::   ||||||
     856 AATGGACTCAGAGAGAAAAAGGCATATTCAAGCTGGTGGATTCAAAGGCT 905

     287 ValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAs 303
         |||:::::|||||||||   :::|||||:::|||   ||||||||:::
     906 GTCTCTAAGCTTTGGGGAAAGCATAAGAACAAACCAGACATGAACTATGA 955

     303 pLysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleIleArgL 320
         :   ::::::|||:::::|||   ||||||:::::|||||:::   |
     956 AACCATGGGACGAGCTTTGAGATACTACTACCAAAGGGGAATTCTTGCAA 1005

     320 ysProAspIleSerGlnArgLeuValTyrGlnPhe 341
         ||   :::   :::|||||||||:|||||||||||
    1006 AGGTTGAA...GGACAGAGGCTTGTATATCAGTTC 1037

seq name  "Sequence 3240 BP; 1070 A; 592 C; 662 G; 916 T; 0 other;" AAC18356
```

```
Seq documentation block:
ID   AAZ18456 standard; cDNA; 5427 BP.
XX
AC   AAZ18456;
XX
DT   19-OCT 1999  (first entry)
XX
DE   Human ASTHIJ alt1 cDNA sequence.
XX
KW   ASTHI; asthma; human; chromosome 11p; ASTHII; ASTHIJ; genetic locus;
KW   therapeutic; immunogen; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             150..1052
FT                   /*tag   a
FT                   /product   "ASTHIJ protein"
XX
PN   WO9933809 A1.
XX
PD   29 JUL 1999.
XX
PF   21-JAN 1998;   98WO US01260.
XX
PR   21 JAN-1998;   98WO US01260.
XX
PA   (AXYS ) AXYS PHARM INC.
XX
PI   Brooks-Wilson AR,  Buckler A,  Cardon L,  Carey AH;
PI   Galvin M,  Miller A,  North M;
XX
DR   WPI; 1999 479058/40.
DR   P PSDB; AAY41416.
XX
PT   Mammalian asthma related genes, useful for diagnosis of a
PT   predisposition to development of asthma
XX
PS   Claim 15; Page 97-99; 196pp; English.
XX
CC   The invention identifies a genetic locus ASTHI, associated with asthma,
CC   mapped to human chromosome 11p. ASTHII and ASTHIJ are genes present
CC   within the locus, located close to each other on human chromosome 11p,
CC   and have similar patterns of expression, and common sequence motifs. The
CC   ASTHI genes and fragments, encoded protein, genomic regulatory regions
CC   and anti ASTHI antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for the modulation of gene
CC   activity in vivo for prophylactic and therapeutic purposes. The ASTHI
CC   protein is useful as an immunogen to raise specific antibodies, in drug
CC   screening for compositions that mimic or modulate ASTHI activity or
CC   expression, including altered forms of ASTHI protein, and as a
CC   therapeutic. The present sequence represents a human ASTHIJ alt1
CC   cDNA sequence.
XX
SQ   Sequence 5427 BP; 1617 A; 1047 C; 1147 G; 1616 T; 0 other;


alignment scores:
            Quality:  262.00              Length:     463
              Ratio:  1.424                 Gaps:      18
 Percent Similarity:  50.689   Percent Identity: 25.895%

alignment block:
 US-09-126-945B-2 x AAZ18456   ..

 Align seq 1/1  to: AAZ18456  from: 1  to: 5427

      41 ArqAspTrpSerProSerProProAlaThrProGluGlnGly..... 54
         |||  |||  |||  |||  ::::::  |||  :::
      88 CGCCCGTGGTGCCCCATCCCTATAGGAGCTGGTGAGATTGCAGCCTGCTG 87

      55 .LeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGlu.... 69
         |||  ::::::     |||||:::  :::  ;::||||||
```

```
      88 CCTCCCCTCCATCAGCCACAGCTATTGG...ATTTCCCACCCAGAATCTT 134

      70 ..............AspSerSer.....TrpAlaAlaLysAlaPro 78
                       |||||||:::     |||:::  :::
     135 TAGGTAAATGAGATCATGATTCTGGAAGGAGGTGGTGTAATGAA..... 178

      79 GlyAlaSerSerArqGluGluProProGluGluProGluGlnCysProVa 95
         ::::::  |||:::::::|||||
     179 ...TCTCAACCCCGGCAACAACCTCCT..................... 202

      96 IleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyL 112
                 :::||||||:::|||||||||
     203 ............TCACCAGCCGCCAGCCTGGAC.............. 223

     113 euThrLeuGluGlu.HisSerLeuGluGlnValGlnSerMetValValGl 128
             :::::: :::|||      :::|||:::|||         ||
     224 .......AGAcAGcTAcTCcACtGcAATGTTCCAGTGGGTTTTTGG 265

     129 yGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrA 145
         |          :::|||:::
     266 AGGCCACTGGCATGAAATCAT..........................  287

     146 laAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrp 161
           ||||::   |||:::      :::|||   :::|||||||
     288 .....CCTCAGTACTGGACCAAGTACCAGGTGTGGGAGTGGCTCCAGCAC 332

     162 ThrGlnHisGluTyrArqLeuProProMetGlyLysAlaPheGluGln.. 177
                 :::|||                 |||||||||
     333 CTCCTGGACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTT 382

     178 ....LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArqG 193
             :::   |||:::::::|||||:::|||||   ::::::::   :
     383 CGACATCAACGGCGAGCACCTCTGCAGCATGAGTTTGCAGGAGTTCACCC 432

     194 lnArqSerProLeuGlyGlyAspValLeuHisAlaHisLeu..... 206
         ::   :::      :::|||:::::::|||:::::::|||
     433 GGGCGGCAGGGACGGCGGGGCAGCTCCTCTACAGCAACTTGCAGCATCTG 482

     207 ........................AspIleTrpLysSer.....Al 212
                                 |||:::::::|||       ::
     483 AAGTGGAACGGCCAGTGCAGTACTGACCTGTTCCAGTCCACACACAATGT 532

     212 aAlaTrpMetLysGluArqThrSerProGlyAlaIle..... 224
         :   ||:::|||:::|||:::   :::
     533 CATTGTCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAG 582

     225 .....HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGlu 239
           :::|||     :::|||:::     |||   :::|||
     583 ACGAGAACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG... 629

     240 ValAspSerSer.....CysSerGlyGln..... 247
         :::|||||:::      |||   :::|||
     630 TTGGACAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAG 679

     248 .......Profile..................................... 249
               |||:::
     680 TCACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCC 729

     250 ..............................................HisLeuTrp 252
                                                  |||||||||
     730 CTGCCAACTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGG 779

     253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArqPheIl 269
         :::|||||||||||||||||||:|||   :::              :::||
     780 GAATTCATCCGCCACATCCTCTTGAACCCAGACAAGAACCCAGCATTAAT 829

     269 eArqTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaG 286
         |::::||   :::::::::::::|||:::|||:::::       |||
     830 AAAATGGGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGG 879

     286 lnValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMetAsnTyr 302
           |||||||:::|||||||||   :::|||||||:::   :::|||:::|||
     880 CAGTGGCTCAGCTATGGGGTAAAAACAAGAACAACAGCAGCATGACCTAT 929
```

```
      303 AspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 349
          :::|||||||||||||||:::::|||  ||||||||||:::  |||:::::
      940 GAAAAGCTCAGCCGAGCTATGACATATTACTACAAAAGAGAAATACTGGA 989

      319 qLysProAspIleSerGlnArqLeuValTyrGlnPhe 353
          ::::  |||  ::::::|||||||||||||:::|||
      980 GCGTGTGGAT...GGACGAAGACTGGTATATAAATTT 1013

Seq name: US10126/c Database: pctext.gb Type: DNA Length: 5427

seq_documentation_block:
ID   AAA80254 standard; cDNA; 5427 BP.
XX
AC   AAA80254;
XX
DT   22-NOV-2000  (first entry)
XX
DE   Human ASTH1J cDNA, splice variant alt1.
XX
KW   ASTH1 locus; ASTH1J; human; chromosome 11p; asthma;
KW   bronchial hyperreactivity; ets family; transcription factor;
KW   splice variant; genetic predisposition; polymorphism; antibody;
KW   drug screening; prophylaxis; therapy; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             150..1052
FT                   /*tag= a
FT                   /product= "Human ASTH1J"
XX
PN   US6087485-A.
XX
PD   11 JUL-2000.
XX
PF   21-JAN-1998;   98US-0009913.
XX
PR   21-JAN-1997;   97US-0035664.
PR   01 JUL-1997;   97US-0051432.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M;  Miller A;  North M;  Cardon L;  Buckler A;
PI   Brooks-Wilson AR;  Carey AH;
XX
DR   WPI; 2000-505109/45.
DR   P-PSDB; AAB11968.
XX
PT   New nucleic acids other than naturally occurring chromosomes encoding
PT   ASTH1 protein, for e.g. screening compositions that modulate expression
PT   or function of ASTH1 proteins or as diagnostics for genetic
PT   predisposition to asthma -
XX
PS   Claim 3; Column 111-116; 141pp; English.
XX
CC   The invention relates to the ASTH1 locus on the short arm of human
CC   chromosome (11p). This locus comprises the ASTH1I and ASTH1J genes,
CC   which are associated with a genetic predisposition to asthma and
CC   bronchial hyperreactivity. The ASTH1I and ASTH1J genes are oriented in
CC   opposite directions with the ASTH1 locus, and have similar patterns of
CC   expression and common sequence motifs. They are both expressed in
CC   trachea, lung and several other tissues. ASTH1I and ASTH1J are novel
CC   members of the ets family of transcription factors, which have been
CC   implicated in the activation of a variety of genes including the TCRa
CC   gene and cytokine genes known to be important in the aetiology of
CC   asthma. Both ASTH1I and ASTH1J mRNAs are alternatively spliced.
CC   Alternative splicing of transcripts has no effect on the open reading
CC   frame of ASTH1J, as the exons involved are all 5' to the start codon in
CC   exon b. In contrast, alternative splicing of ASTH1I transcripts results
CC   in 3 different ASTH1I isoforms. The invention also encompasses mouse
CC   asth1j protein. The ASTH1 nucleic acids are useful as diagnostics to
CC   identify a hereditary predisposition to asthma, as probes for identifying
CC   ASTH1 related genes, for identifying expression of the gene in a
CC   biological specimen, and for generating genetically modified non-human
CC   animals or site specific gene modifications in cell lines. The encoded
CC   ASTH1 proteins are useful as immunogens to raise specific antibodies; in
CC   drug screening for compositions that mimic or modulate activity or
CC   expression of ASTH1I and/or ASTH1J (including altered forms of these
CC   proteins); and as a therapeutic. The ASTH1 genes or fragments thereof,
CC   encoded proteins, ASTH1 genomic regulatory regions, and anti-ASTH1I and
CC   anti-ASTH1J antibodies are useful in the identification of individuals
CC   pre-disposed to development of asthma and for modulation of gene activity
CC   in vivo for prophylactic and therapeutic purposes. The intact ASTH1I or
CC   ASTH1J proteins or active fragments thereof may be used to modulate or
CC   reduce bronchial hyperreactivity. Sequences AAA80254 A80256 represent
CC   three alternatively spliced human ASTH1J cDNAs.
XX
SQ   Sequence 5427 BP; 1617 A; 1047 C; 1147 G; 1616 T; 0 other;


alignment_scores:
            Quality:  262.00              Length:    363
              Ratio:  1.424                 Gaps:     18
 Percent Similarity:  50.689   Percent Identity: 25.895

alignment_block:
 US-09-126-945B-2 x AAA80254   ..

 Align seq 1/1  to: AAA80254  from: 1  to: 5427

      41 ArqAspIrpSerProSerProProAlaThrProGluGlnGly..... 54
         |||   |||   |||   |||   ::::::   |||   :::
      38 CGCCCGTGGTGCCCCATCCCTATAGGAGCTGGTGAGATTGCAGCCTGCTG 87

      55 .LeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGlu.... 69
          |||   ::::::      ||||||||::   :::   :::||||||
      88 CCTCCCCTCCATCAGCCACAGCTATTGG...ATTTCCCACCCAGAATCTT 134

      70 ................AspSerSer......TrpAlaAlaLysAlaPro 78
                         |||||||::      ||||::   :::
     135 TAGGTAAATGAGATCATGATTCTGGAAGGAGGTGGTGTAATGAA...... 178

      79 GlyAlaSerSerArqGluGluProProGluGluProGluGlnCysProVa 95
          ::::::   |||:::::::|||||
     179 ...TCTCAACCCCGGCAACAACCTCCT..................... 202

      95 IIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyL 112
                 :::|||||::::|||||||||
     203 ............TCACCAGCCGGCCAGCCTGGAC.............. 223

     112 euThrLeuGluGlu.HisSerLeuGluGlnValGlnSerMetValValGl 128
             :::::: :::|||      :::|||:::|||         ||
     224 ........AGACAGCTACTCCACCTCCAATCTTCCACTCCCTTTTTCCC 265

     128 yGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrA 145
         |        :::|||:::
     266 AGGCCAGTGGCATGAAATTCAT........................... 287

     145 laAspProMetAspIrpSerProSerAsnValGlnLysIrpLeuLeuIrp 161
           ||:::   |||:::      :::|||   :::|||||
     288 ....CCTCAGTACTGGACCAAGTACCAGGTGTGGGAGTGGCTCCAGCAC 332

     162 ThrGluHisGlnTyrArqLeuProProMetGlyLysAlaPheGlnGlu.. 177
                   :::|||                    |||||||||
     333 CTCCTGGACACCAACCAGCTGGATGCCAAATTGTATCCCTTTCCAAGAGTT 382

     178 ....LeuAlaGlyLysGluLeuLysAlaMetSerGluGluGlnPheArqG 193
             :::   |||::::::|||||||:::||||||   ::::::|||   :
     383 CGACATCAACGGCGAGCACCTCTGCAGCATGAGTTTGCAGGAGTTCACCC 432

     193 lnArqSerProLeuGlyGlyAspValLeuHisAlaHisLeu......... 206
         ::   :::      :::|||::::::|||::::::::|||
     433 GGGCGGCAGGGACGGCGGGGCAGCTCCTCTACAGCAACTTGCAGCATCTG 482

     207 ......................AspIleTrpLysSer.........Al 212
```

```
                                  |||:::::::::::|||          ::
     483 AAGTGGAACGGCCAGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGT 532

     212 aAlaTrpMetLysGluArgThrSerProGlyAlaIle..............  234
         :          ||||:|||:::|||||:   :::
     533 CATTGTCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAG 582

     225 .....HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGlu  239
              :::|||      :::|||:::      |||   :::|||
     583 ACGAGAACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG...  629

     240 ValAspSerSer......CysSerGlyGln.....................  247
         :::||||||:::       |||   :::|||
     630 TTGGACAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAG 679

     248 .......Profile.....................................  249
                 ||||::
     680 TCACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCC 729

     250 .....................................HisLeuTrp  252
                                              |||||||||
     730 CTGCCAAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGG 779

     253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl  269
         :::|||:::::::::::::|||||:::|||   :::         :::||
     780 GAATTCATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAAT 829

     269 cArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaG  286
         |:::|||   :::::::::::|||:::|||:::::       |||
     830 AAAATGGGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGG 879

     286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnGly  302
           ||||||:::|||||||||   :::||||||:::   :::|||:::|||
     880 CAGTGGCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCIAI 929

     303 AspLysLeuSerArgSerIleArgGlyTyrTyrLysLysGlyIleIleAr  319
         :::|||||||||||:::::|||   ||||||||||:::   |||::::::
     930 GAAAAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATACTGGA 979

     319 qLysProAspIleSerGlnArgLeuValTyrGlnPhe  331
         ::::   |||   ::::::|||||||||||||:::|||
     980 GCGTGTGGAT...GGACGAAGACTGGTATATAAATTT 1013

Seq name: gi|18357|pir|A28652|seq_json pg_NAL_CC_DAT_AA_1.CC

Seq documentation block:
ID   AAZ18357 standard; cDNA; 5510 BP.
XX
AC   AAZ18357;
XX
DT   19-OCT-1999  (first entry)
XX
DE   Human ASTHIJ alt2 cDNA sequence.
XX
KW   ASTH1; asthma; human; chromosome 11p; ASTH11; ASTHIJ; genetic locus;
KW   therapeutic; immunogen; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             232..1135
FT                   /*tag= a
FT                   /product= "ASTHIJ protein"
XX
PN   WO9937809-A1.
XX
PD   29-JUL-1999.
XX
PF   21-JAN-1998;   98WO-US01260.
XX
PR   21-JAN-1998;   98WO-US01260.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Brooks-Wilson AR;  Buckler A;  Cardon L;  Carey AH;
PI   Galvin M;  Miller A;  North M;
XX
DR   WPI; 1999-479058/40.
DR   P-PSDB; AAY41416.
XX
PT   Mammalian asthma related genes, useful for diagnosis of a
PT   predisposition to development of asthma
XX
PS   Claim 15; Page 99-101; 195pp; English.
XX
CC   The invention identifies a genetic locus ASTH1, associated with asthma,
CC   mapped to human chromosome 11p. ASTH11 and ASTHIJ are genes present
CC   within the locus, located close to each other on human chromosome 11p,
CC   and have similar patterns of expression, and common sequence motifs. The
CC   ASTH1 genes and fragments, encoded protein, genomic regulatory regions
CC   and anti-ASTH1 antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for the modulation of gene
CC   activity in vivo for prophylactic and therapeutic purposes. The ASTH1
CC   protein is useful as an immunogen to raise specific antibodies, in drug
CC   screening for compositions that mimic or modulate ASTH1 activity or
CC   expression, including altered forms of ASTH1 protein, and as a
CC   therapeutic. The present sequence represents a human ASTHIJ alt2
CC   cDNA sequence.
XX
SQ   Sequence 5510 BP; 1624 A; 1064 C; 1190 G; 1641 T; 1 other;


Alignment scores:
            Quality:  262.00              Length:     329
              Ratio:   1.492                Gaps:      14
 Percent Similarity:  53.191   Percent Identity: 25.228

Alignment block:
US-09-126-945B-2 x AAZ18357  ..

Align seq 1/1  to: AAZ18357  from: 1  to: 5510

     71 SerSerTrpAlaAlaLysAlaProGlyAlaSerSer......ArgGluGl  85
        |||:::|||:::::::::::::   |||::::::        :::   ||
    152 TCAGCTTGGGTAAGCCGGGTCTGCGGCGGGGATTGCCATCTGAAGACAGA 201

     86 GC Tr:Le:Ar:Al:Al:Ly:Al:Pr:Gl:Al:Se:Se:Ar:Va:Le:As:Se:Gl:Al:Pr:A  102
        |          :::|||:::   :::   ::::::      ::::::   :
    202 GGCAGGAGGGCAGCCACACCTTGCCCAGATCATGATTCTGGAAGGA...G 248

    102 laGlySerLeuAspLeuValProGlyGlyLeuThrLeu.............  114
        ::|||   :::::::|||   |||||:::      |||
    249 GTGGGTGTAATGAATCTCAACCCCGGCAACAACCTCCTCACCAGCCGCCA 298

    115 .........GluGluHisSerLeuGluGluValGlnSerMetValValGl  128
                 ::::::::::|||      :::|||:::|||         ||
    299 GCCTGGACAGACAGCTACTCCACGTGCAATGTTTCCAGTGGGTTTTTGG 348

    128 yGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrA  145
        |          :::|||:::
    349 AGGCCAGTGGCATGAAATTCAT.............................. 370

    145 laAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrp  161
             |||:::   |||:::      :::|||   :::||||||
    371 .....CCTCAGTACTGGACCAACTACCAGGTGTGGGACTGGCTCCAGCAC 415

    162 ThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGluGlu..  177
                    :::|||                  ||||||||
    416 CTCCTGGACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTT 465

    178 ....LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgG  193
            :::   |||::::::|||||||:::|||||   :::::::|||   :
    466 CGACATCAACGGCAGCACCTCTGCAGCATCAGTTTGCAGGAGTTCACCC 515

    193 lnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu.........  206
        ::   :::      :::|||:::::::|||:::::::::|||
```

```
      516 GGGAGGCAGGGACGGCGGGGCAGCTCCTCTACAGCAACTTGCAGCATCTG 565

      217 .....AspIleTrpLysSer.....Al 232
                                 |||:::::::::::|||          ::
      566 AAGTGGAAAGGGCAGTGCAGTAGTGACCTGTTCCAGTCCACAC A AA1GT 615

      233 aAlaTrpMetLysGluArgThrSerProGlyAlaIle..... 224
          :            |||:::|||:::|||:::   :::
      616 CATTGTCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAG 665

      225 .....HisTyrCysAlaSerThrSerGlyGlnSerTrpThrAspSerGl 274
            :::|||      :::|||:::      |||   :::|||
      666 ACGAGAACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG... 712

      240 ValAspSerSer.....CysSerGlyGln..... 247
          :::||||||:::     |||   :::|||
      713 TTGGACAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAG 762

      248 .......ProIle..... 249
                |||:::
      763 TCACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCC 812

      250 .....HisLeuTrp 252
                        ||||:|||
      813 CTGCAAGTGCCACACCAAAAAGCACAACCCGAGAGGCACTCACTTATGG 862

      253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPro 299
          :::|||||||||||||||||||:|||   :::     :::||
      863 GAATTCATCCTGCAGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAAT 912

      269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGlnAspCysAla 318
          :::|||   :::::::::::|||:::|||:::      |||
      913 AAAATGGGAAGACGGATCTGAGGGCGTCTTCAGTTCTTGAAATCAGAGG 962

      286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
            |||||:::|||||||||   :::|||||:::   :::|||:::|||
      963 CAGTGGCTCAGCTATGGGCTAAAAAGAAGAACAACAGCAGCATGACCTAT 1012

      303 AspLysLeuSerArgSerIleArgIleTyrTyrLysLysGlyIleIleAr 319
          :::|||||||||||||:::|||   |||||||||||:   |||::::::
     1013 GAAAAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATACTGGA 1062

      319 gLysProAspIleSerGlnArgLeuValTyrGlnPhe 351
          ::::   |||   :::::::|||||||||||||:|||
     1063 GCGTCTGCAT...GGACGAAGACTGGTATATAAATTT 1096

seq_name: /cgi-bin/gcgdata/genseq/genseq/NA2000_1AT AAA80255

seq_documentation_block:
ID   AAA80255 standard; cDNA; 5510 BP.
XX
AC   AAA80255;
XX
DT   22-NOV-2000  (first entry)
XX
DE   Human ASTH1J cDNA, splice variant alt2.
XX
KW   ASTH1 locus; ASTH1J; human; chromosome 11p; asthma;
KW   bronchial hyperreactivity; ets family; transcription factor;
KW   splice variant; genetic predisposition; polymorphism; antibody;
KW   drug screening; prophylaxis; therapy; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             234..1145
FT                   /*tag= a
FT                   /product= "Human ASTH1J"
XX
PN   US6087485-A.
XX
PD   11-JUL-2000.
XX
PF   21-JAN-1998;   98US-0009913.
XX
PR   21-JAN-1997;   97US-0035663.
PR   01-JUL-1997;   97US-0051432.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  North M,  Cardon L,  Buckler A;
PI   Brooks-Wilson AR,  Carey AH;
XX
DR   WPI, 2000-505107/45.
DR   P-PSDB; AAB11968.
XX
PT   New nucleic acids other than naturally occurring chromosomes encoding
PT   ASTH1 protein, for e.g. screening compositions that modulate expression
PT   or function of ASTH1 proteins or as diagnostics for genetic
PT   predisposition to asthma -
XX
PS   Claim 4; Column 117-122; 131pp; English.
XX
CC   The invention relates to the ASTH1 locus on the short arm of human
CC   chromosome (11p). This locus comprises the ASTH1I and ASTH1J genes,
CC   which are associated with a genetic predisposition to asthma and
CC   bronchial hyperreactivity. The ASTH1I and ASTH1J genes are oriented in
CC   opposite directions with the ASTH1 locus, and have similar patterns of
CC   expression and common sequence motifs. They are both expressed in
CC   trachea, lung and several other tissues. ASTH1I and ASTH1J are novel
CC   members of the ets family of transcription factors, which have been
CC   implicated in the activation of a variety of genes including the ICAM
CC   gene and cytokine genes known to be important in the aetiology of
CC   asthma. Both ASTH1I and ASTH1J mRNAs are alternatively spliced.
CC   Alternative splicing of transcripts has no effect on the open reading
CC   frame of ASTH1J, as the exons involved are all 5' to the start codon in
CC   exon b. In contrast, alternative splicing of ASTH1I transcripts results
CC   in 3 different ASTH1I isoforms. The invention also encompasses mouse
CC   asth1j protein. The ASTH1 nucleic acids are useful as diagnostics to
CC   identify a hereditary predisposition to asthma, as probes for identifying
CC   ASTH1 related genes, for identifying expression of the gene in a
CC   biological specimen, and for generating genetically modified non-human
CC   animals or site specific gene modifications in cell lines. The encoded
CC   ASTH1 proteins are useful as immunogens to raise specific antibodies; in
CC   drug screening for compositions that mimic or modulate activity or
CC   expression of ASTH1 and/or ASTH1J (including altered forms of these
CC   proteins); and as a therapeutic. The ASTH1 genes or fragments thereof,
CC   encoded proteins, ASTH1 genomic regulatory regions, and anti ASTH1 and
CC   anti-ASTH1J antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for modulation of gene activity
CC   in vivo for prophylactic and therapeutic purposes. The intact ASTH1I or
CC   ASTH1J proteins or active fragments thereof may be used to modulate or
CC   reduce bronchial hyperreactivity. Sequences AAA80254 AAA80256 represent
CC   three alternatively spliced human ASTH1J cDNAs.
XX
SQ   Sequence 5510 BP; 1624 A; 1064 C; 1191 G; 1631 T; 0 other;

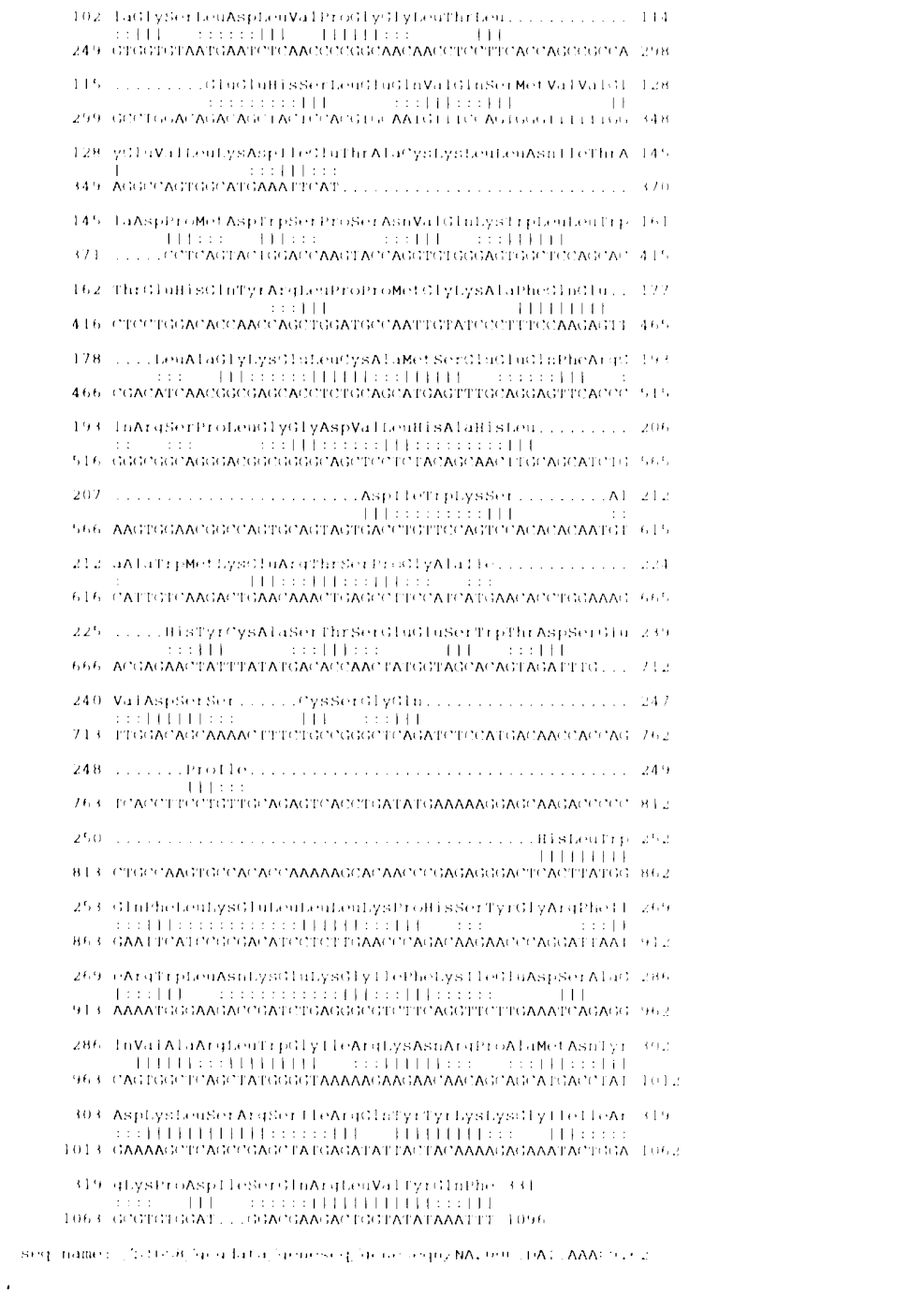

alignment_scores:
            Quality:  262.00              Length:    329
              Ratio:  1.497                 Gaps:     13
 Percent Similarity:  53.191   Percent Identity: 25.228

alignment_block:
 US-09-126-945B-2 x AAA80255   ..

 Align seq 1/1  to: AAA80255  from: 1  to: 5510

      71 SerSerIrpAlaAlaLysAlaProGlyAlaSerSer.....ArgGluGl 85
         |||:::|||:::::::::::   |||::::::          :::   ||
     152 TCAGCTTGGGTAAGCCGGGTCTGCGGCGGGGATGCCATCTGAAGACAGA 201

      85 uProProGluGluProGluGlnCysProValIleAspSerGlnAlaProA 102
         |        :::|||:::   :::   ::::::      ::::::   :
     202 GGCAGGAGGGCAGCCACACCTGCCAGATCATGATTCTGGAAGGA...G 248
```



```
seq documentation block:
ID   AAA09162 standard; DNA; 4159 BP.
XX
AC   AAA09162;
XX
DT   10-AUG-2000 (first entry)
XX
DE   Human 36P1A6 gene, a homologue of murine EHF gene.
XX
KW   36P1A6; transcription factor; murine EHF homologue; ETS family;
KW   cytostatic; cancer; vaccine; tumorigenesis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             112..1014
FT                   /*tag= a
FT   misc_signal     109..115
FT                   /*tag= b
FT                   /note= "putative Kozak sequence"
XX
PN   WO200020584-A2.
XX
PD   13-APR-2000.
XX
PF   02-OCT-1999;    99WO-US22576.
XX
PR   02-OCT-1998;    98US-0102744.
PR   29-JUL-1999;    99US-0146447.
XX
PA   (UROG ) UROGENESYS INC.
PA   (AFAR/) AFAR D E.
PA   (HUBE/) HUBERT R S.
PA   (MITC/) MITCHELL S C.
XX
PI   Afar DE, Hubert RS, Mitchell SC;
XX
DR   WPI; 2000-293772/26.
DR   P-PSDB; AAY92243.
XX
PT   Novel putative transcription factor gene 36P1A6 for treatment,
PT   diagnosis and prevention of prostate, bladder, cervical, ovarian,
PT   pancreatic, and colonic cancer
XX
PS   Claim 4; Fig 1A; 53pp; English.
XX
CC   The human 36P1A6 gene encodes a putative transcription factor based on
CC   homology to the murine EHF gene which encodes a transcription factor
CC   which is a member of the ETS family. 36P1A6 is expressed in
CC   androgen-dependent and androgen independent LAPC prostate cancer
CC   xenografts and in normal prostate at approximately equal levels. The
CC   highest expression is in the prostate and colon. 36P1A6 may be involved
CC   in activating tumor promoting genes or repressing genes that block
CC   tumorigenesis. The 36P1A6 polynucleotides and polypeptides are used for
CC   the treatment and diagnosis of cancer, e.g. prostate, bladder, cervical,
CC   ovarian, pancreatic and colonic cancer (all claimed). Anti 36P1A6
CC   antibodies may be used for purifying 36P1A6 and for isolating 36P1A6
CC   homologues. Antisense oligonucleotides and ribozymes can be used to
CC   inhibit the transcription and translation of the 36P1A6 gene (claimed).
CC   The 36P1A6 polynucleotides and polypeptides and immunogenic fragments
CC   may also be used in cancer vaccines (claimed).
XX
SQ   Sequence 4159 BP; 942 A; 633 C; 683 G; 838 T; 63 other;


alignment scores:
            Quality:  258.00              Length:     861
              Ratio:  1.410                 Gaps:      18
 Percent Similarity:  50.693   Percent Identity:  25.762

alignment block:
 sq: 09-126-945b-2 x AAA09162   ..
```

```
Align seq 1/1  to: AAA09162  from: 1  to: 3159

     43 IrpSerProSerProProAlaThrProGluGlnGly.....LeuSe  56
        ||| ||| ||| ||| ::::::  ||| ::: ||| |||
      6 TGGTGCCCCATCCCTATAGGAGCTGGTGAGATTGCAGCCTGCTGCCTCCC  55

     56 rAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGlu..... 69
        :::::::      |||||||||  :::  :::|||||
     56 CTCCATCAGCCACAGCTATTGG...ATTTCCCACCCAGAATCTTTAGGTA 102

     70 ..........AspSerSer......TrpAlaAlaLysAlaProGlyAla 80
                ||||||||:::      ||||:   :::         :::
    103 AATGAGATCATGATTCTGGAAGGAGGTGGTGTAATGAA.....TCT 143

     81 SerSerArgGluGluProProGluGluProGluGlnCysProValIleAs 97
        :::  |||:::::::|||||
    144 CAACCCCGGCAACAACCTCCT..........................  164

     97 pSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrL 114
              :::|||||:::|||||||||
    165 .......TCACCAGCCGCCAGCCTGGAC.....................  185

    114 euGluGln.HisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
        :::::: ::|||      :::|||:::|||       |||
    186 ..AGACAGCTACTCCACGTGCAATGTTCCAGTGGGTTTTTGGAGGCCA 233

    130 lLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspI 147
              :::|||:::                              |
    234 GTGGCATGAAATTCAT..............................C  250

    147 roMetAspIrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGlu 163
        !::::  |||:::      :::|||  :::|||||
    251 CTCAGTACTGGACCAAGTACCAGGTGTGGGAGTGGCTCCAGCACCTCCTG 300

    164 HisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu......Le 178
              :::|||                  |||||||||       ::
    301 GACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTTCGACAT 350

    178 sAlaGlyLysGluLeuCysAlaMetSerGluGlnGlnPheArgGluArgG 195
        :  |||:::::|||||:::|||||  :::::::|||  :::  :
    351 CAACGGCGAGCACCTCTGCAGCATGAGTTTGCAGGAGTTCACCCGGGCGG 400

    195 erProLeuGlyGlyAspValLeuHisAlaHisLeu............... 206
        ::      :::|||:::::::|||:::::::::|||
    401 CAGGGACGGCGGGGCAGCTCCTCTACAGCAACTTGCAGCATCTGAAGTGG 450

    207 ..................AspIleIrpLysSer.........AlaAlaIr 214
                           |||:::::::::::::         :::
    451 AACGGCCAGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGTCATTGT 500

    214 pMetLysGluArgThrSerProGlyAlaIle..................H 229
           |||:::|||:::|||:::   :::                    :
    501 CAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAGACGACA 550

    225 isTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluValAsp 241
        ::|||      :::|||:::      |||   ::::||      :::|||
    551 ACTATTATATGACACCAACTATGGTAGCACAGTAGATTTG...ITGGAC 597

    242 SerSer......CysSerGlyGln.......................... 247
        (|:::      |||   ::::||
    598 AGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAGTCACCT 647

    248 .ProIle.............................................. 249
         ||:::
    648 TCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCCCTGCCA 697

    250 ....................................HisLeuIrpGlnPhe 264
                                           ||||||||||:::|||
    698 AGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGGAATTC 747

    265 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgIs 281
        ::::::::::::|||||:::|||   :::          :::||||:::||
```

```
    748 ATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATAAAATG 797

    271 pLeuAsnLysGluLysGlyIlePheLysIleGlnAspSerAlaGlnValA 288
        |  ::::::::::::::|||:::|||:::::::      |||      ||||
    798 GGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGCAGTGG 847

    288 laArgLeuIrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLys 304
        ||:::|||||||||   :::|||||||:::   :::|||:::|||:|||
    848 CTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTATGAAAAG 897

    305 LeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPr 321
        ||||||||||:::::|||   |||||||||:::   |||:::::::::
    898 CTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATTCTGGAGCGTGT 947

    321 oAspIleSerGlnArgLeuValTyrGlnPhe 331
          |||   ::::::|||||||||||:::|||
    948 GCAT...GGACGAAGACTGGTATATAAATTT 975

seq_name: /SIS/d_serial/prodseq/prodseq/NA2000.DAT:AAC99081

seq_documentation_block:
ID   AAC99081 standard; cDNA; 1426 BP.
XX
AC   AAC99081;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:309.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54316.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition
XX
PS   Claim 1; Page 735; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
```

```
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 1426 BP; 441 A; 331 C; 327 G; 334 T; 3 other;


alignment scores:
            Quality:  255.50              Length:     462
              Ratio:  1.496                 Gaps:      18
 Percent Similarity:  50.552   Percent Identity: 25.691

alignment block:
 US-09-126-945B-2 x AAC99081   ..

 Align seq 1/1  to: AAC99081  from: 1  to: 1426

       41 ArgAspTrpSerProSerProProAlaThrProGluGlnGly..... 54
          |||  |||  |||  |||  ||||||  |||  :::
       83 GGCCCGTGGTGCCCCATCCCTATGGAGCTGGTGAGATGCAGCCTGCTG 82

       55 .LeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGln.... 69
          |||  ||||||     ||||||||:::   :::  ||||||||
       83 CCTCCCCTCCATCAGCCACAGCTATTGG...ATTTCCCACCCAGAATCTT 129

       70 ..................AspSerSer......TrpAlaAlaLysAlaPro 78
                           ||||||||:::      |||:::  :::
      130 TAGGTAAATGAGATCATGATTCTGGAAGGAGCTGGTGTAATGAA...... 173

       79 GlyAlaSerSerArgGluGluProProGluGluProGluGlnCysProVa 95
          ||||||  |||||||||||||||
      174 ...TCTCAACCCCGGCAACAACCTCCT...................... 197

       95 IleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyL 112
                  :::||||||:::|||||||||
      198 ............TCACCAGCCGCCAGCCTGGAC................ 218

      112 ouThrLeuGluGlu.HisSerLeuGluGlnValGlnSerMetValValGl 128
                ||||||  ||||||      ||||||||||||          ||
      219 ........ACACAGCTACTCCACGTGCAATGTTTCCAGTGGGTTTTTTGG 260

      128 yGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleGluA 145
          |           :::|||:::
      261 AGGCCAGTGGCATGAAATTCAT........................... 282

      145 laAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrp 161
               |||:::   |||:::      ::::|||   ::::||||||
      283 ....CCTCAGTACTGGACCAACTACCAGGTGTGGGAGTGGCTCCAGAA 327

      162 ThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu.. 177
                     :::|||                 ||||||||||
      328 CTCCTGCACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGACTT 377

      178 ....LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgG 193
              :::   |||:::::||||||:::||||||   ::::::|||   :
      378 CGACATCAACGGCGAGCACCTTGCAGCATGAGTTTGCAGGAGTTCACCC 427

      193 InArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu........ 206
          ::   :::      :::|||:::::::|||:::::::::|||
      428 GGGCGGCAGGGACGGCGGGGCAGCTCCTCTACAGCAACTTGCAGCATCTG 477

      207 ........................AspIleTrpLysSer.........Al 212
                                  |||:::::::::|||        ::
      478 AAGTGGAACGGCATGCACGTAGTGACCTGTTCCAGTCCACACACAATGT 527

      212 aAlaTrpMetLysGluArgThrSerProGlyAlaIle............. 224
          :            |||:::|||:::|||:::   :::

      528 CATTGTCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAG 577

      225 .....HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGlu 239
               :::|||      :::||||:::      |||   ::::||
      578 ACCAGAACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG... 624

      240 ValAspSerSer......CysSerGlyGln..................... 247
          :::||||||||:      |||   ::::||
      625 TTGGACAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAG 674

      248 .......Profile.................................... 249
                ||||::
      675 TCACCTTCCTGTTGAGTCACCTGATATGAAAAACGAGCAAGACCCCCCTG 724

      250 .....................................HisLeuTrpGln 253
                                              ||||||||||::
      725 CCAAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGGAA 774

      254 PheLeuLysGluGnLeuLeuLysProHisSerTyrGlyArgProIleAr 270
          |||:::::::::::::||||:::|||   :::          :::||:::
      775 TTCATCCGGGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATAAA 824

      270 qTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnV 287
          :|||   :::::::::::::|||:::|||:::::      |||      |
      825 ATGGGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGCAG 874

      287 alAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAsp 303
          |||||:::|||||||||   :::|||  :::  :::|||:::|||:::
      875 TGGCTCAGCTATGGGGTAAAAAGAAGAMCAACAGCAGCATGACCTATGAA 924

      304 LysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLy 320
          ||||||||||||:::::::|||   ||||||||||:::   |||:::::::::
      925 AAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATTCTGGAGCG 974

      320 sProAspIleSerGluArgLeuValTyrGluPhe 331
          :   |||   ::::::||||||||||||||:::|||
      975 TGTGGAT...GCACGAAGACTGGTATATAAATTT 1005

Seq [illegible]

seq documentation block:
ID   AAC77813 standard; cDNA; 1996 BP.
XX
AC   AAC77813;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:207.
XX
KW   human; cancer associated gene, cancer antigen, detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective, thrombolytic, coagulant; nootropic;
KW   vasotropic; antipsoriatic; antianqiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05882.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
```

```
XX
DR   WPI; 2000-587533/55.
DR   P-PSDB; AAB43604.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer -
XX
PS   Claim 1; Page 780-781; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given
CC   in AAB43398 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective, cardiant, thrombolytic, coagulant,
CC   nootropic; vasotropic; antipsoriatic and antianqiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 1996 BP; 472 A; 575 C; 556 G; 391 T; 2 other;


alignment_scores:
            Quality:  255.00              Length:    435
              Ratio:  1.371                 Gaps:     17
 Percent Similarity:  42.759   Percent Identity: 22.989

alignment_block:
 US-09-126-945B-2 x AAC77813   ..

 Align seq 1/1  to: AAC77813  from: 1  to: 1996

      25 ArqThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArqAr 41
         ||||:|||      :::|||||||||      ::::::          ||
      24 AGAGCCGGGTAGGGGAGCGCAGCRGCCAGATACCTCAGCGCTACCTGGCG 73

      41 qAspTrpSerProSerProPro.......................... 48
         |:::|||         ||||||
      74 GAACTGGATTTCTCTCCCGCCTGCCGGCCTGCCTGCCACAGCCGGACTCC 123

      49 .....................AlaThrProGluGlnGlyLeuSerAla 57
                              ||||||   |||      :::|||
     124 GCCACTCCGCTAGCCTCATGGCTGCAACCTGTCAG.....ATTAGCAAC 167

      58 PheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAl 74
         ::::::   :::||||||||:   :::|||   ::::::::|||
     168 ATTTTTAGCAACTACTTCAGTGCGATGTACAGCTCGGAGGACTCCACCCT 217

      74 aAlaLysAlaProGlyAlaSerSerArqGluGluProProGluGlnProG 91
          |||:::::::|||   |||
     218 GGCCTCTGTTCCCCCTGCT.............................. 236

      91 TuGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeu 107
                                 |||     |||:::   ||||||
     237 ............................GCCACCTTTGGGGCCGATGACTTG 260

     108 ValProGlyGlyLeuThrLeuGlu......GluHisSerLeuGluGlnVa 122
         |||         ||||||||||:      :::   |||||||||   ::
     261 GTA.........CTGACCCTGAGCAACCCCCAGATGTCATTGGAGGGTAC 301

     122 tGln...SerMetValValGlyGluValLeuLysAspIleGluThrAlaC 138

         ::::   :::      :::|||||||
     302 AGAGAAGGCCAGCTGGTTGGGGGAA.......................... 326

     138 ysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsn 154
                                 :::|||:::   ||||||   ::::::
     327 ......................CAGCCCCAGTTCTGGTCGAAGACGCAG 353

     155 ValGluLysTrpLeuLeuTrp......ThrGlnHisGlnTyrArqLeuPr 169
         |||      ||||::   :::          ::::::::::|||
     354 GTTCTGGACTGGATCAGCTACCAAGTGGAGAAGAACAAGTACGACGCAAG 403

     169 oProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaM 186
           :::      :::      ::::::   |||      ||||||
     404 CGCCATTGACTTCTCACGATGTGACATGGATGGCGCCACCCTCTGCAATT 453

     186 etSerGluGluGlnPheArq...GlnArqSerProLeuGlyGlyAspVal 201
           :::   |||:::::::|||         :::|||||   ||||||
     454 GTGCCCTTGAGGAGCTGCGTCTGGTCTTTGGGCCTCTG   GGGGACCAA 500

     202 LeuHisAlaHisLeu...AspIleTrpLysSerAla.............Al 213
         ||||||||||:::|||   ||:::   :::|||:::            ::
     501 CTCCATGCCCAGCTGCGAGACCTCAC TTCCAGCTCTTCTGATGAGCTCAG 550

     213 aTrpMet.................................LysGluArqT 219
         :||||:                                :::|||
     551 TTGGATCATTGAGCTGCTGGAGAAGGATGGCATGGCCTTCCAGGAGGCCC 600

     219 hrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGlnSerTrp 235
         :::|||||   :::      :::      :::::::
     601 TAGACCCAGGGCCCTTGACCAGGGCAGCCCCTTGCCCAGGAGCTGCTG 650

     236 ThrAspSerGluValAspSer..............SerCysSer..... 245
           |||::::::      |||              |||||:::
     651 GACGACGGTCAGCAAGCCAGCCCCTACCACCCCGGCAGCTGTGGCGCAGG 700

     245 ..................................................... 245

     701 AGCCCCCTCCCCYGGCAGCTCTGACGTCTCCACCGCACGGACTGCTGCTT 750

     245 ..................................................... 245

     751 CTCGGAGCTCCCACTCCTCAGACTCCGGTGGAAGTGACGTGGACCTGGAT 800

     245 ..................................................... 245

     801 CCCACTGATGGCAAGCTCTTCCCCAGCGATGGTTTTCGTGACTGCAAGAA 850

     246 ..............................GlyGlnPro.......... 248
                                     |||::::::
     851 GGGGGATCCCAAGCACGGGAAGCGGAAACGAGGCCGGCCCCGAAAGCTGA 900

     248 ..................................................... 248

     901 GCAAAGAGTACTGGGACTGTCTCGAGGGCAAGAAGAGCAAGCACGCGCCC 950

     249 ......IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHi 268
               |||||||||:::|||:::::::::::::::::::::   ||:::
     951 AGAGGCACCCACCTGTGGGACTTCATCGGGGACATCCTCATCCACCCGGA 1000

     268 sSerTyrGlyArqPheIleArqTrpLeuAsnLysGluLysGlyIlePheL 280
         :          :::::::::|||   |||:::::::::|||::::|||
    1001 GCTCAACGAGGGCCTCATGAAGTGGAGAATCGGCATGAAGGCGTCTTCA 1050

     280 ysIleGluAspSerAlaGlnValAlaArqLeuTrpGlyIleArqLysAsn 296
         |:::      |||      |||||:::|||||||   :::|||:::
    1051 AGTTCCTGCGCTCCGAGGCTGTGGCCCAACTATGGGGCCAAAAGAAAAAG 1100

     297 ArqProAlaMetAsnTyrAspLysLeuSerArqSerIleArqGlnTyrTy 313
         :::      |||:::|||:::|||||||||||:::::|||   |||||
    1101 AACAGCAACATGACCTACGACAAGCTGAGCCGGGCCATGAGGCTACTACTA 1150

     313 rLysLysGlyIleIleArqLysProAspIleSerGlnArqLeuValTyrG 330
         |||:::   |||:::::::::   |||   ::::::|||||||||||:
```

```
    1151 CAAACGGCAGATCCTGGAACGGGTGCAT...GGCCGGCGACTCGTCTACA 1197

     330 InPhe 341
         ::|||
    1198 AGTTT 1202

Seq name: /S1DS8/gcgdata/geneseq/geneseqn/NA1998_DAT:AAV20465

Seq documentation block:
ID   AAV20465 standard; DNA; 1604 BP.
XX
AC   AAV20465;
XX
DT   17-JUN-1998  (first entry)
XX
DE   Human c-ets oncogene.
XX
KW   Human; oncogene; proto oncogene; neoplastic disease; anticancer;
KW   cancer; antisense oligonucleotide; c-ets; ds.
XX
OS   Homo sapiens.
XX
PN   US5734049-A.
XX
PD   31-MAR-1998.
XX
PF   15-SEP-1994;   94US-0306691.
XX
PR   15-SEP-1994;   94US-0306691.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Calabretta B,  Skorski T;
XX
DR   WPI; 1998-229882/20.
XX
PT   Anticancer composition comprising two anti-sense oligonucleotide(s)
PT       targetting cytoplasmic and nuclear oncogene(s)
XX
PS   Claim 1; Column 125-128; 92pp; English.
XX
CC   The present sequence represents an oncogene from the present invention.
CC   The present invention describes a composition which comprises two
CC   antisense oligonucleotides. The first oligonucleotide is specific for a
CC   cytoplasmic oncogene or proto-oncogene selected from ras, raf, EGF-1,
CC   c-fms, c-ros, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-tgr and
CC   c-yes. The second oligonucleotide is specific for a nuclear oncogene or
CC   proto-oncogene selected from myc, jun, c-ets, c-fos, c-myb, B-myb,
CC   c-rel, c-vav, c-ski, c-api, cyclin-D1, PML/RAR-alpha, AML1/MTG8,
CC   E2A/pC1 and ALL-1/AF-4. The composition is used for treating cancer.
CC   The combination of antisense oligonucleotides has synergistically
CC   enhanced ability to inhibit growth of cancer cells.
XX
SQ   Sequence 1604 BP; 418 A; 443 C; 408 G; 335 T; 0 other;


alignment scores:
            Quality:  251.50              Length:     540
              Ratio:  1.251                 Gaps:      22
 Percent Similarity:  37.925   Percent Identity: 21.698

alignment block:
 US-09-126-945B-2 x AAV20465   ..

 Align seq 1/1  to: AAV20465  from: 1  to: 1604

       24 SerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluAr 40
          :::|||:::|||   |||:::   |||   |||:::
       10 GCGAGGGCCGGGCAGGAGGAGCGGGCGCGCGGCGGG.............. 45

       40 qArqAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerA 57
          |||   |||:::||||||      :::::|||:::::::::
       46 CGCAGGCTGGGACCCGAGCGCGCTCACTTCGCCGCAAAGTGCC....... 87

       57 laPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrp 73
                                                :::      |||
       88 .....................................AACTTCCCCTGG 99

       74 AlaAla...LysAlaProGly..AlaSerSerArgGluGluPro..... 86
          ...|||   ...||||||...  |||   .....:|||:::
      100 AGTGCCGGGCGCGCACCGTCCGGGCGCGGGGGAAAGAAAGGCAGCGGGAA 149

       87 ............................ProGluGluProGluGlnCy 93
                                       |||   :::|||:::   ||
      150 TTTGAGATTTTTGGGAAGAAAGTCGGATTTCCCCCGTCCCCTTCCCCCTG 199

       93 s.............................................P 94
          |                                            |
      200 TTACTAATCCTCATTAAAAAGAAAAACAACAATAACTGCAAACTTGCTAC 249

       94 roValIleAspSerGlnAlaProAla.........GlySerLeuAspLeu 107
          ||          |||   :::|||:::         :::::::::||||||
      250 CATCCCGTACGTCCCCCACTCCTGGCACCATGAAGGCGGCCGTCGATCTC 299

      108 ValPro..................................... 109
          |||
      300 AAGCCGACTCTCACCATCATCAAGACGGAAAAAGTCGATCTGGAGCTTTT 349

      110 .....................GlyGlyLeuThrLeuGluGluHisSerL 119
                              :::   :::   |||         |||
      350 CCCCTCCCCGGATATGGAATGTGCAGATGTCCCACTATTAACTCCAAGCA 399

      119 euGluGlnValGlnSerMetValValGlyGluValLeuLysAspIleGlu 135
          :::::::::::||||:::::::::     :::::::::   :::
      400 GCAAAGAAATGATGTCTCAAGCATTAAAAGCTACTTTCAGTGGTTTCACT 449

      136 ThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerPr 152
                  :::   |||:::|||      ||||||   :::|||:::
      450 AAAGAACAGCAACGACTGGGGATCCCAAAAGACCCCCGGCAGTGGACAGA 499

      152 oSerAsnValGluLysTrpLeuLeuTrpThrGluHisGluTyrArgLeuP 169
          ::::::|||:::   |||:::::::|||:::   :::::::::   |||
      500 AACCCATGTTCGGGACTGGGTGATGTGGGCTGTGAATGAATTCAGCCTG. 548

      169 roProMetGlyLysAlaPheGlnGlu......LeuAlaGlyLysGluLeu 183
                |||       ||||||:::      :::   |||     |||
      549 .....AAAGGTGTAGACTTCCAGAAGTTCTGTATGAATGGAGCAGCCCTC 593

      184 CysAlaMetSerGluGluGlnPheArgGlnArgSerPro...LeuGlyGl 199
          ||||||:::::::::::   |||   :::   :::|||   :::   ||
      594 TGCGCCCTGGGTAAAGACTGCTTTCTCGAGCTGGCCCCAGACTTTGTTGG 643

      199 yAspValLeuHisAlaHisLeuAspIleTrpLys.............. 210
          ||||:::|||      ||||||:::|||   :::
      644 GGACATCTTATGGGAACATCTAGAGATCCTGCAGAAAGAGGATGTGAAAC 693

      211 ...................SerAlaAlaTrpMetLysGluArq...Thr 219
                             :::   |||:::   ::::::|||   |||
      694 CATATCAAGTTAATGGAGTCAACCCAGCCTATCCAGAATCCCGCTATACC 743

      220 SerProGlyAlaIleHisTyr.................CysAlaSerTh 230
          |||         |||   |||                 |||:::
      744 TCGGATTACTTCATTAGCTATGGTATTGAGCATGCCCAGTGTGTTCCACC 793

      230 rSerGlu.......................GluSerTrp.......... 235
          ||||||                      |||||||:::
      794 ATCGGAGTTCTCAGAGCCCAGCTTCATCACAGAGTCCTATCAGACGCTCC 843

      235 ..................................................... 235

      844 ATCCCATCAGCTCGGAACAGCTCCTCTTCCCTCAAGTATGAGAATGACTAC 893

      236 ............................ThrAspSer............ 238
                                       |||||||:::
      894 CCCTCGGTCATTCTCCGAGACCCTCTCCAGACAGACACCTTGCAGAATGA 943
```

```
 239 ..................GluVal.........AspSerSerCysSerG 216
                       ||||||         |||:::   |||   |
 944 CTACTTTGGTATCAAACAAGAAGTCGTCACCCCAGACAACATGTGCATCG 893

 246 lyGln..................................................... 247
     ||:::
 994 GGAGGACCAGTCGTGGTAAACTCGGGGGCCAGGACTCTTTTGAAAGCATA 1043

 247 ..................................................... 247

1044 GAGAGCTACGATAGTTGTGATCGCCTCACCCAGTCCTGGAGCAGCCAGTC 1093

 247 ..................................................... 247

1094 ATCTTTCAACAGCCTGCAGCGTGTTCCCTCCTATGACAGCTTCGACTCAG 1143

 247 ..................................................... 247

1144 AGGACTATCCGGCTGCCCTGCCCAACCACAAGCCCAAGGGCACCTTCAAG 1193

 247                                                       247

1194 GACTATGTGCCGGACCGTGCTGACCTCAATAAGGACAAGCCTGTCATTCC 1243

 248 .......................        ProIleHisLeuTrpG 253
                                    |||||||:||||||
1244 TGCTGCTGCCCTACCTGCCTACACAGGCAGTGGACCAATCCAGCTATGGC 1293

 253 laPheLeuLysGluLeuLeuLysProHisSerTyrGlyArgPheIle 269
     ||||||||   |||   ||||||         |||         ||||||
1294 AGTTTCTTCTGGAA   TTACTGAGCACATAAATCCTCTCAGTCTTTATC 1340

 270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 286
       |||   :::   :::         ||||||:::::|||      ::
1341 AGCTGGACAGGA   GATGGCTGGGAATTCAAACTTTCTGACCCAGATGA 1387

 286 nValAlaArgLeuTrpGlyIleArgLysAsnArgPheAlaMetAsnTyrA 302
     :|||||||||   ||||||   ||||||||||:::|||   |||||||||:
1388 GGTGGCCAGGAGATGGGGAAAGAGGAAAAACAAACCTAAGATGAATTATG 1437

 303 spLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArg 319
     ::|||||||||||||:::::|||   ||||||   |||:::||||||:::
1438 AGAAACTGAGCCGTGGCCTACGCTACTATTACCACAAAAACATCATCCAC 1487

 320 LysProAspIleSerGlnArgLeuValTyrGlnPheVal 332
     |||         :::::|||   ||||||:::||||||
1488 AAG...ACAGCGGGGAAACGCTACGTGTACCGCTTTGTG 1523

seq_name /SIDS4/gcgdata/genseqp/genseqp/NA1998.DAT:AAV40701

seq_documentation_block:
ID   AAV40701 standard; DNA; 1920 BP.
XX
AC   AAV40701;
XX
DT   21-SEP-1998 (first entry)
XX
DE   Human ELF3 protein coding sequence.
XX
KW   ELF3; human; cancer; epithelial cancer; breast cancer; prostate cancer;
KW   diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             115..1230
FT                   /*tag= a
FT                   /product= ELF3
XX
PN   EP849908-A2
XX
PD   06-MAY-1998.
XX
PF   23-OCT-1997;    97EP-0308450.
XX
PR   15-NOV-1996;    96US-0746789.
PR   31-OCT-1996;    96US-0028791.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (MONU ) UNIV MONASH.
XX
PI   Debouck C, Kola I, Tymms MJ;
XX
DR   WPI; 1998-212574/22.
DR   P-PSDB; AAW57834.
XX
PT   New isolated polynucleotide and e.g. DNA, agonists and antagonists -
PT   used to diagnose and treat cancer, e.g. prostate, breast, lung or
PT   epithelial tumours
XX
PS   Claim 4; Fig 1A; 34pp; English.
XX
CC   This sequence encodes the human ELF3 protein of the invention. ELF3 is
CC   used to treat patients requiring ELF3 by providing DNA encoding the
CC   polypeptide and expressing the polypeptide in vivo. ELF3 is used to
CC   diagnose a disease or a susceptibility to a disease related to its
CC   expression by determining a mutation in the nucleic acid sequence
CC   encoding it. The presence ELF3 in a sample derived from a host can be
CC   detected. Antagonists of ELF3 are used to treat patients requiring
CC   inhibition of ELF3 polypeptide. ELF3 can be used in the early diagnosis
CC   and treatment of cancer, particularly epithelial cancer such as breast
CC   and prostate cancer.
XX
SQ   Sequence 1920 BP; 472 A; 561 C; 550 G; 387 T; 0 other;


alignment_scores:
            Quality:  250.00              Length:    429
              Ratio:  1.374                 Gaps:     17
 Percent Similarity:  42.424   Percent Identity: 22.844

alignment_block:
 US-09-126-945B-2 x AAV40701   ..

 Align seq 1/1  to: AAV40701  from: 1  to: 1920

      31 AlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPr 47
         |||||||||      ::::::       |||:::|||         ||
      16 GCAGCGGCCAGATACCTCAGCGCIACCTGGCGGAACTGGATTTCTCTCCC 65

      47 oPro..................................................... 48
         ||||
      66 GCCTGCCGGCCTGCCTGCCACAGCCGGACTCCGCCACCCCGGTAGCCTCA 115

      49 .....AlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPhe 63
              ||||||   |||       :::|||   ::::::   :::||||||
     116 TGGCTGCAACCTGTGAG......ATTAGCAACATTTTTAGCAACTACTTC 159

      64 AspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAl 80
         :::   :::|||   ::::::::::|||       |||::::::::|||   ||
     160 AGTGCCATGTACAGCTCGGAGGACTCCACCTTGGCCTCTGTTCCCCCTGC 209

      80 aSerSerArgGluGluProProGluGluProGluGlnCysProValIleA 97
         |
     210 T..................................................... 210

      97 spSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThr 113
                |||       |||:::   |||||||||          ||||||
     211         GCCACCTTTGGGGCCGATGACTTGCTA.........CTGACC 243

     114 LeuGlu......GluHisSerLeuGluGlnValGln...SerMetValVa 127
         |||:::      :::   |||||||||   ::::::   :::      ::
     241 CTGAGCAACCCCCAGATCTCATTGGAGCGTACAGACAACGCTAGCTGGTT 290

     127 lGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleT 144
         :||||||
```

```
294 GGGGGAA..........                                                  300

144 hrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeu 160
      :::||||:::   ||||||  :::::::|||     |||:::
301 .....CAGCCCCAGTTCTGGTCCAAGACCCAGGTTCTGGACTGGATCAGC 345

161 Trp......ThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPh 175
    :::      :::::::::|||           :::      :::
346 TACCAAGTGGAGAAGAACAAGTACGACCCAAGCGCCATTGACTTCTCACG 395

175 eGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheA 192
      ::::::   |||     ||||||     :::   ||||||||
396 ATGTGACATGGATGGGCCCACCCTCTGCAATTGTGCCCTTGAGGACCTGC 445

192 rq...GlnArqSerProLeuGlyGlyAspValLeuHisAlaHisLeu... 206
    ||       :::|||||   ||||||  ||||||||||:::|||
446 GTCTGCTCTTTGGGCCTCTG...GGGCACCAACTCCATGCCCAGCTGCGA 492

207 AspIleTrpLysSerAla..........AlaTrpMet........... 219
    |||:::   :::|||:::          :::|||:::
493 GACCTCACTTCCAGCTCTTCTGATGAGCTCAGTTGGATCATTGAGCTGCT 542

216 ....................LysGluArqThrSerProGlyAlaIleH 225
                        :::|||     :::||||||   :::
543 GGAGAAGGATGGCATGGCCTTCCAGGAGGCCCTAGACCCAGGGCCCTTG 592

225 isTyrCysAlaSerThrSerGluGlaSerTrpThrAspSerGluValAsp 241
          :::     :::::::|||        |||::::::
593 ACCAGGGCAGCCCCTTGCCCAGGAGCTGCTGGACGACGGTCAGCAAGCC 642

242 Ser..............SerCysSer..................... 245
    |||              ||||||:::
643 AGCCCCTACCACCCCGGCAGCTGTGGCGCAGGAGCCCCTCCCCTGGCAG 692

245 ..................................................... 245

693 CTCTGACGTCTCCACCGCAGGGACTGGTGCTTCTCGGAGCTCCCACTCCT 742

245 ..................................................... 245

743 CAGACTCCCGTGGAAGTGACGTGGACCTGGATCCCACTGATGGCAAGCTC 792

245 ..................................................... 245

793 TTCCCCAGCGATGGTTTTCGTGACTGCAAGAAGGGGGATCCCAAGCACGG 842

246 ............GlyGlnPro........................... 248
                |||:::|||
843 GAAGCGGAAACGAGGCCGGCCCCGAAAGCTGAGCAAAGAGTACTGGGACT 892

249 ....................................IleHisLeuTrp 252
                                        |||||||||
893 GTCTCGAGGGCAAGAAGAGCAAGCACGCGCCCAGAGGCACCCACCTGTGG 942

253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArqPheH 269
    :::|||:::::::::::::|||:::   |||:::           :::::
943 GAGTTCATCCGGGACATCCTCATCCACCCGGAGCTCAACGAGGGCCTCAT 992

269 eArqTrpLeuAsnLysGlnLysGlyIlePheLysIleGlnAspSerAlaG 286
    :::|||   |||:::::::::|||:::::|||||:::      |||
993 GAAGTGGGAGAATCGGCATGAAGGCGTCTTCAAGTTCCTGCGCTCCGAGG 1042

286 lnValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMetAsnTyr 302
      ||||||:::|||||||||   :::|||:::::::      |||:::|||
1043 CTGTGGCCCAACTATGGGGGCAAAAGAAAAGAACAGCAACATGACCTAC 1092

303 AspLysLeuSerArqSerIleArqGlnTyrTyrLysLysGlyIleIleAr 319
    :::|||||||||||||:::::|||   ||||||||||:::   |||:::::
1093 GAGAAGCTGAGCCGGGCCATGAGGTACTACTACAAACGGGAGATCCTGGA 1142

319 qLysProAspIleSerGlnArqLeuValTyrGlnPhe 341
    ::::   |||   :::::|||||||||||||:::|||
1143 ACCGGTGGAT...GGCCGGCGACTCGTCTACAAGTTT 1176
```

```
...i_orf ... 1..1907 /db_xref="pid:g4680998.DAI:AAW48355"

Seq_documentation block:
ID   AAW48355 standard; cDNA; 1907 BP.
XX
AC   AAW48355;
XX
DT   09-NOV-1998 (first entry)
XX
DE   Transcription factor ESX cDNA (ORF and untranslated regions).
XX
KW   ESX; epithelial restricted with serine box; transcription factor;
KW   ETS; human; epithelial cancer; breast cancer; adenocarcinoma;
KW   diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..95
FT                   /*tag= a
FT   CDS             96..1206
FT                   /*tag= b
FT                   /note= "Claim 4"
FT   3'UTR           1207..1907
FT                   /*tag= c
XX
PN   WO9823782-A2.
XX
PD   04-JUN-1998.
XX
PF   26-NOV-1997;    97WO-US21865.
XX
PR   25-NOV-1997;    97US-0031504.
PR   27-NOV-1996;    96US-0031504.
XX
PA   (REGC-) UNIV CALIFORNIA.
XX
PI   Benz CC, Chang C, Scott GK;
XX
DR   WPI; 1998-322755/28.
DR   P-PSDB; AAW60677.
XX
PT   Nucleic acid encoding the ESX transcription activator -
PT   over-expressed in epithelial, especially breast, cancers, also
PT   related polypeptide(s), antibodies, vectors and transformed cells,
PT   useful for diagnosis and treatment of cancer
XX
PS   Disclosure; Page 83; 120pp; English.
XX
CC   This cDNA sequence comprises the open reading frame (see also
CC   AAV48354) and 5' and 3' untranslated regions of the newly isolated
CC   ETX gene that codes for a novel human transcription factor (see
CC   AAW60677) associated with the aetiology of cancers. The ETX gene
CC   is a member of the ETS family and is designated ETX for epithelial
CC   restricted with serine box. The gene is located at chromosome
CC   1q42, a region amplified in 50% of early breast cancers. ESX is an
CC   early indicator of breast cancer, being expressed at the ductal
CC   carcinoma in situ stage. Dysfunction of the ESX gene that results
CC   in increased expression is indicative of epithelial cancer
CC   (especially breast cancer but also gastric, ovarian and lung
CC   adenocarcinoma) and of unfavourable prognosis. The ESX open
CC   reading frame was deduced from a cDNA clone of foetal liver spleen
CC   origin that was identified in an expressed sequence tag database
CC   search using human Ets 2, and from a clone obtained by 5'RACE of
CC   placental polyA mRNA. Antibodies, antisense nucleic acids or
CC   inactive ESX mutant proteins that inhibit activity of ESX can be
CC   used to inhibit growth and proliferation of neoplastic cells,
CC   particularly cancer. ESX polynucleotides, polypeptides and
CC   antibodies are also used to detect ESX, particularly for diagnosis
CC   and monitoring of cancer. ESX polynucleotides may also be used to
CC   detect gross abnormalities in chromosome 1, e.g. for in utero
CC   screening for predisposition to cancer, and to generate ESX protein
```

```
CC   in vivo from gene therapy vectors.  ESX protein, genes and DNA can
CC   also be used to screen for specific modulators, especially
CC   potential anticancer agents.
XX
SQ   Sequence 1907 BP; 427 A; 557 C; 533 G; 390 T; 0 other;


alignment_scores:
            Quality:  246.00              Length:    419
              Ratio:  1.390                 Gaps:     17
 Percent Similarity:  42.243   Percent Identity: 22.673

alignment_block.
 US-09-126-945B-2 x AAV38355   ..

 Align seq 1/1  to: AAV38355  from: 1  to: 1907

      41 ArgAspTrpSerProSerProPro..........................  48
         |||:::|||      ||||||
      27 CGGAACTGGATTTCTCTCCCGCCTGCCGGCCTGCCTCCCACAGCCGGACT  76

      49 .........................AlaThrProGluGlnGlyLeuSerA  57
                                  ||||||   |||      :::|||
      77 CCGCCACTCCGGTAGCCTCATGGCTGCAACCTGTGAG......ATTAGCA 120

      57 laPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrp  73
           ::::::   :::|||||||:::   :::|||   ::::::::::|||
     121 ACATTTTTAGCAACTACTTCAGTGCGATGTACAGCTCGGAGGACTCCACC 170

      74 AlaAlaLysAlaProGlyAlaSerSerArgGluGluProProGluGluPr  90
           |||::::::|||   |||
     171 CTGGCCTCTGTTCCCCCTGCT............................. 191

      90 oGluGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspL 107
                                  |||      ||||::   ||||
     192 ..........................GCCACCTTTGGGGCCGATGACT 213

     107 euValProGlyGlyLeuThrLeuGln      GluHisSerLeuGluGln 121
         |||||         ||||||||||:::      :::   |||||||||
     214 TGGTA.........CTGACCCTGAGCAACCCCCAGATGTCATTGGAGGGT 254

     122 ValGln...SerMetValValGlyGlnValLeuLysAspIleGluThrAl 137
         ::::::   :::      :::|||||
     255 ACAGAGAAGGCCAGCTGGTTGGGGGAA..................... 281

     137 aCysLysLeuLeuAsnIlethrAlaAspProMetAspTrpSerProSerA 154
                                :::|||:::   ||||||   ::::
     282                        CAGCCCCAGTTCTGGTCGAAGACGC 306

     154 snValGlnLysTrpLeuLeuTrp......ThrGluHisGlnTyrArgLeu 168
         :::||      ||||::   :::         ::::::::::|||
     307 AGGTTCTGGACTGGATCAGCTACCAAGTGGACAAGAACAAGTACGACGCA 356

     169 ProProMetGlyLysAlaPheGluGluLeuAlaGlyLysGluLeuCysAl 185
             :::      :::      ::::::   |||      ||||||
     357 AGCGCCATTGACTTCTCACGATGTGACATGGATGGCGCCACCCTCTGCAA 406

     185 aMetSerGluGluGlnPheArg...GluArgSerProLeuGlyGlyAspV 201
         :::   |||:::::|||         :::|||||   ||||||
     407 TTGTGCCCTTGAGGAGCTGCGTCTGGTCTTTGGGCCTCTG...GGGGACC 453

     201 alLeuHisAlaHisLeu...AspIleTrpLysSerAla............. 212
         ||||||||||::|||   |||:::   :::|||:::
     454 AACTCCATGCCCAGCTGCGAGACCTCAGTGCCAGCTCTTCTGATGAGGTC 503

     213 AlaTrpMet................................LysGluAr 218
         :::|||:::                               ::::||
     504 AGTTGGATCATTGAGCTGCTGGACAAGGATGGCATGGCTTTCCAGGAGGC 553

     218 qThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerT 235
           ::|||||   :::      :::      :::::|||
     554 CCTAGACCCAGGGCCCTTTGACCAGGGCAGCCCCTTTGCCCAGGAGCTGC 603

     235 rpThrAspSerGluValAspSer.........        SerCysSer  245
           |||:::::      |||                   |||||||:::
     604 TGGACCACGGTCAGCAAGCCAGCCCCTACCACCCCGGCAGCTGTGGCGCA 653

     245                                       ..............  245

     654 GGAGCCCCCTCCCCTGGCAGCTCTGACGTCTCCACCGCAGGGACTGGTGC 703

     245 ..................................................  245

     704 TTCTCGGAGCTCCCACTCCTCAGACTCCGGTGGAAGTGACGTGGACCTGG 753

     245 ..................................................  245

     754 ATCCCACTGATGGCAAGCTCTTCCCCAGCGATGGTTTTCGTGACTGCAAG 803

     246 ..............................GlyGlnPro.........  248
                                        |||:::|||
     804 AAGGGGGATCCCAAGCACGGGAAGCGGAAACGAGGCCGGCCCCGAAAGCT 853

     248 ..................................................  248

     854 GAGCAAACAGTACTGGGACTGCTCCGAGGGCAAGAAGAGCAAGCACGGC 903

     249 ........IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPro 262
                 ||||||||||:::|||:::::::::::::|||:::   |||
     904 CCAGAGGCACCCACCTGTGGGAGTTCATCCGGGACATCCTCATCCACCCG 953

     263 HisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePh 279
         :::            ::::::::::|||   |||::::::::::|||:::||
     954 GAGCTCAACGAGGGCTCATGAAGTGGGAGAATCGGCATGAAGGCGTCTT 1003

     279 eLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysA 296
         ||||:::      |||      |||||||:::||||||||   :::|||:
    1004 CAAGTTCCTGCGCTCCGAGGCTGTGGCCCAACTATGGGGCCAAAAGAAAA 1053

     296 snArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyr 312
         :::::      |||:::|||:::|||||||||||||:::::|||   |||
    1054 AGAACAGCAACATGACCTACGAGAAGCTGAGCCGGGCCATGAGGTACTAC 1103

     313 TyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTy 329
         |||||||::   |||::::::::::   |||   :::::|||||||||||
    1104 TACAAACGGGAGATCCTGGAACGGGTGGAT...GGCCGGCGACTCGTCTA 1150

     329 rGlnPhe 331
         |:::|||
    1151 CAAGTTT 1157

Seq_name: /SIDS8/gcgdata/gcgseqsq/genseqbp/NA2000.DAT:AA236707

seq_documentation_block:
ID   AA236707 standard; DNA; 1907 BP.
XX
AC   AA236707;
XX
DT   13-MAR-2000  (first entry)
XX
DE   Nucleotide sequence for GenBank accession number 866804.
XX
KW   Stimulus-regulated nucleic acid; sequence profile; nucleic acid level;
KW   differentially expressed nucleic acid; disease state; cancer;
KW   autoimmune disease; infectious disease; aging; developmental disorder;
KW   proliferative disorder; neurological disorder; toxicity;
KW   treatment resistance; differential expression, drug discovery;
KW   growth factor; epidermal growth factor; radiation; stress; pathogen; ss.
XX
OS   Homo sapiens.
XX
PN   WO9955913-A2.
XX
PD   04-NOV-1999.
XX
PF   27-APR-1999;   99WO-US09119.
```

```
XX
PR   27 APR 1998;   98US 0083431.
PR   27 AUG 1998;   98US 0098070.
PR   04 FEB 1999;   99US 0118624.
XX
PA   (KIMM ) KIMMEL CANCER CENT SIDNEY.
XX
PI   McClelland M,  Welsh J,  Trenkle T;
XX
DR   WPI; 2000-086388/07.
XX
PT   Measuring expression of low abundance reduced complexity target nucleic
PT   acid molecules
XX
PS   Disclosure; Fig 34; 187pp; English.
XX
CC   AAZ36681-Z36725 represent nucleotide sequences from stimulus-regulated
CC   nucleic acid molecules. The sequences represent a profile of sequences
CC   which can function as targets in the method of the invention. The
CC   specification describes a method for measuring the level of two or more
CC   nucleic acid molecules in a target. The method comprises contacting a
CC   probe with an arbitrarily or statistically sampled target and detecting
CC   the amount of specific binding of the target to the probe. The methods
CC   can be used to identify differentially expressed nucleic acid molecules
CC   associated with disease states, such as cancer, autoimmune disease,
CC   infectious disease, aging, developmental disorder, proliferative
CC   disorder or neurological disorder. Alternatively the methods can be
CC   used to assess the efficacy or toxicity of or a resistance to a
CC   treatment. Also the methods can be used to determine differential
CC   expression of nucleic acid molecules in response to a stimulus, e.g. a
CC   chemical, drug or growth factor (especially epidermal growth factor),
CC   radiation, stress or a pathogen. The methods can also be used to
CC   determine co-regulated genes that can be potential targets for drug
CC   discovery.
XX
SQ   Sequence 1907 BP; 427 A; 557 C; 534 G; 390 T; 0 other;


Alignment scores:
            Quality:  246.00              Length:    419
              Ratio:  1.390                 Gaps:     17
 Percent Similarity:  42.243   Percent Identity: 22.673

Alignment block:
 US-09-126-945B-2 x AAZ36707   ..

 Align seq 1/1  to: AAZ36707  from: 1  to: 1907

       41 ArgAspTrpSerProSerProPro..........................  48
          |||:::|||          ||||||
       27 CGGAACTGGATTTCTCTCCCGCCTGCCGGCCTGCCTGCCACAGCCGGACT  76

       49 .........................AlaThrProGluGlnGlyLeuSerA  57
                                  ||||||   |||       :::|||
       77 CCGCCACTCCGGTAGCCTCATGGCTGCAACCTGTGAG.....ATTAGCA  120

       57 IaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrp  73
          ::::::   :::||||||:::   :::|||   ::::::::|||
      121 ACATTTTTAGCAACTACTTCAGTGCGATGTACAGCTCGGAGGACTCCACC  170

       74 AlaAlaLysAlaProGlyAlaSerSerArgGluGluProProGluGluPr  90
             |||::::::|||   |||
      171 CTGGCCTCTGTTCCCCCTGCT.............................  191

       90 oGluGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspL  107
                                 |||      |||:::   ||||
      192 ..........................GCCACCTTTGGGGCCGATGACT  213

      107 euValProGlyGlyLeuThrLeuGlu.....GluHisSerLeuGluGlu  121
          |||||         ||||||||||:::       :::   |||||||||
      214 TGGTA.........CTGACCCTGAGCAACCCCCAGATGTCATTGCAGGGT  254

      122 ValGln...SerMetValValGlyGluValLeuLysAspIleGluThrAl  137
          ::::::   :::      :::||||||
      255 ACAGAGAAGGCCAGCTGGTTGGGGGAA.......................  281

      137 aCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerA  154
                                  :::||:::   ||||||   ::::
      282 .........................CAGCCCCAGTTCTGGTCGAAGACGC  306

      154 snValGluLysTrpLeuLeuTrp......ThrGluHisGluTyrArgLeu  168
          ::|||       ||:::   :::          ::::::::::|||
      307 AGGTTCTGGACTGGATCAGCTACCAAGTGGACAAGAACAAGTACGACGCA  356

      169 ProProMetGlyLysAlaPheGluGluLeuAlaGlyLysGluLeuCysAl  185
             :::      :::     ::::::   |||      ||||||
      357 AGCGCCATTGACTTCTCACGATGTGACATGGATGGCGCCACCCTCTGCAA  406

      185 aMetSerGluGluGlnPheArg...GlnArgSerProLeuGlyGlyAspV  201
          :::   |||:::::::||         :::||||||   ||||||
      407 TTGTGCCCTTGAGGAGCTGCGTCTGGTCTTTGGGCCTCTG...GGGGACC  453

      201 alLeuHisAlaHisLeu...AspIleTrpLysSerAla.............  212
            ||||||||||::|||   |||:::   :::|||:::
      454 AACTCCATGCCCAGCTGCGAGACCTCACTTCCAGCTCTTCTGATGAGCTC  503

      213 AlaTrpMet.....................................LysGluAr  218
          :::|||:::                                   ::|||
      504 AGTTGGATCATTGAGCTGCTGGAGAAGGATGGCATGGCCTTCCAGGAGGC  553

      218 gThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerT  235
          :::||||||   :::       :::      ::::::::|||
      554 CCTAGACCCAGGGCCCTTGACCAGGGCAGCCCCTTTGCCCAGGAGCTGC  603

      235 rpThrAspSerGluValAspSer..............SerCysSer....  245
          |||::::::       |||                 |||||:::
      604 TGGACGACGGTCAGCAAAGCCAGCCCCTACCACCCCGGCAGCTGTGGCCA  653

      245 .....................................................  245

      654 GGAGCCCCCTCCCCTGGCAGCTCTGACGTCTCCACCGCAGGGACTGGTGC  703

      245 .....................................................  245

      704 TTCTCGGAGCTCCCACTCCTCAGACTCCGGTGGAAGTGACGTGGAGCTGG  753

      245 .....................................................  245

      754 ATCCCACTGATGGCAAGCTCTTCCCCAGCGATGGTTTTCGTGACTGCAAG  803

      246 ..................................GlyGlnPro.........  248
                                           |||::|||
      804 AAGGGGCATCCCAAGCACGGGAAGCGGAAACGAGGCCGGCCCCGAAAGCT  853

      248 .....................................................  248

      854 GAGCAAAGAGTACTGGACTGTCTCGAGGGCAAGAAGAGCAAGCACGCGC  903

      249 ........IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPro  262
                  ||||||||||:::|||||||||||||:::::   |||
      904 CCAGAGCACCCACCTGTGGGAGTTCATCCGGGACATCCTCATCCACCCG  953

      263 HisSerTyrGlyArgPheIleArgTrpLeuArgLysGluLysGlyIlePh  229
          :::          ::::::::|||   |||:::::::::::|||||
      954 GACCTCAACGAGGGCCTCATGAAGTGGCAGAATCGGCATGAAGGCGTCTT  1003

      279 eLysIleGluAspSerAlaGluValAlaArgLeuTrpGlyIleArgLysA  296
          ||||:::      |||      |||||||:::||||||||   :::|||:
     1004 CAAGATCCTGCGCTCCGAGGCTGTGGCCCAACTATGGGGCCAAAAGAAAA  1053

      296 snArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyr  312
          :::::      |||:::|||:::||||||||||||:::::|||   |||
     1054 AGAACAGCAACATGACCTACGAGAAGCTGAGCCGGCCCATGAGGTACTAC  1104

      313 TyrLysLysGlyIleIleArgLysProAspIleSerGluArgLeuValTy  329
          ||||||:::   |||:::::::::   |||   :::::||||||||||
```

```
     1104 TACAAACGGGAGATCCTGGAACGGGTGGAT...GGCCGGCGACTCGTCTA 1153

      329 rGlnPhe 331
          ||:|||
     1151 CAAGTTT 1157

Seq_name: /cgi68/geniatk/pubseq/pubseqdb/NA1998.DAT.AAV15529

Seq_documentation_block:
ID   AAV15529 standard; cDNA; 2975 BP.
XX
AC   AAV15529;
XX
DT   22-MAY-1998  (first entry)
XX
DE   Human lymphoid-specific transcription factor NERF-1 cDNA.
XX
KW   ets-related; human; lymphoid-specific transcription factor; NERF-1;
KW   screening assay; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             122..1687
FT                   /*tag= a
FT                   /product= "NERF-1"
XX
PN   US5721113-A.
XX
PD   24-FEB-1998.
XX
PF   03-JAN-1995;   95US-0468281.
XX
PR   03-JAN-1995;   95US-0468281.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Kunsch CA;  Libermann TA;  Oettgen JP;
XX
DR   WPI; 1998-168405/15.
DR   P-PSDB; AAW47237.
XX
PT   NERF-1 and NERF-2 nucleic acids  encode ets related human
PT   lymphoid-specific transcription factors
XX
PS   Claim 1; Columns 17-22; 17pp; English.
XX
CC   The present sequence encodes the ets-related lymphoid specific
CC   transcription factor NERF-1, which can be used in screening assays
CC   for drugs that modulate NERF activity, and to treat patients having
CC   need of NERF-1. ets-related factors are involved in, e.g. cancer
CC   development, retrovirus replication, T-cell cytokine production,
CC   cell cycle effects, growth regulation and cell differentiation.
XX
SQ   Sequence 2975 BP; 958 A; 575 C; 614 G; 828 T; 0 other;


alignment_scores:
            Quality:  242.00              Length:     261
              Ratio:  1.847                 Gaps:       6
 Percent Similarity:  58.192   Percent Identity:  25.287

alignment_block:
 US-09-126-945B-2 x AAV15529   ..

 Align seq 1/1  to: AAV15529  from: 1  to: 2975

      99 GlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGl 115
         :::::||:::::::::|||||||||:::      ::::::      ::
     140 GAGGGACCCACGAACCAGCTGGATCTGCTCATCCGGGCCGTGGAAGCATC 189

     115 uGluHisSer.....................LeuGluGlnValG 128
         :   ||||||                     ::::|   ::::

     190 AGTTCACAGCAGTAATGCACACTGTACAGATAAGACAATTGAAGCTGCTG 239

     128 lnSerMetValValGlyGlu.........ValLeuLysAsp..... 148
         ::::::::::::       |||           |||:::|||
     240 AAGCCCTGCTTCATATGGAATCTCCTACCTGCTTGAGGGATTCAAGAAGT 289

     144 .....................................IleGluThrAl 162
                                              |||||||::
     290 CCTGAATTCATCCATGCTGCTATGAGGCCAGATGTCATTACAGAAACTGT 339

     163 aCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerA 154
         :  ::::::   :::      ::::::|rrrrrrr   ||||||
     340 AGTGGAGGTGTCAACTGAAGAGTCTGAACCCATGGATACCTCTGCT.... 385

     154 snValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProPro 170
                                                 ::::||
     386 .....................................ATTCCAACA 394

     171 MetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSe 187
                 :::   :::   :::      |||:::          :::::
     395 TCACCAGATAGCCATGAACCAATGAAAAAGAAAAAA.........GTTGC 435

     187 rGluGluGlnrheArgGlnArgSerProLeuGlyGlyAspValLeuHisA 204
         :::::::          |||:::||||||:::::::::
     436 CCGTAAACCAAAGACCCAGCAATCACCAATTTCCAATGGG.......... 475

     204 laHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSer 220
                       :::          ::::|||:::::
     476 ................TCTCCTGAGTTAGGTATAAAGAAGAAACCAAGA 508

     221 ProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAs 237
           |||
     509 GAAGGA.................................................. 514

     237 pSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnP 254
                                  :::|||:::      :::||||||:::
     515 .....................AAAGGAAACACAACCTATTTGTGGGAGT 542

     254 heLeuLysGlnLeuLeuLeuLysProHisSerTyrGlyArgPheIleArg 270
         |||||   :::||||||         ::::::     |||:::|||:::
     543 TTCTTTAGATCTACTTCAAGATAAAAAATACTTGTCCCAGGTATATTAAA 592

     271 TrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGluVa 287
         |||   :::::|||||||||||||||||||:::   ||||||      ||
     593 TGGACTCAGAGAGAAAAAGGCATATTCAAGCTGGTGGATTCAAAGGCTGT 642

     287 lAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspL 304
         |:::::|||||||||   :::|||||:::|||   ||||||||||:::
     643 CTCTAAGCTTTGGGGAAAGCATAAGAACAAACCAGACATGAACTATGAAA 692

     304 ysLeuSerArgSerIleArgLeuTyrTyrLysLysGlyIleIleArgLys 320
         :::::|||:::::|||   ||||||:::::|||||||:::   |||
     693 CCATGGGACGAGCTTTGAGATACTACTACCAAAGGGGAATTCTTGCAAAG 742

     321 ProAspIleSerGlnArgLeuValTyrGluPhe 331
         :::   :::|||||||||||||||||||||||
     743 GTTGAA...GGACAGAGGCTTGTATATCAGTTC 772

Seq_name: /cgi68/geniatk/pubseq/pubseqdb/NA1999.DAT.AAZ18688

Seq_documentation_block:
ID   AAZ18688 standard; cDNA; 848 BP.
XX
AC   AAZ18688;
XX
DT   19-OCT-1999  (first entry)
XX
DE   Mouse asthlj cDNA sequence.
XX
KW   AsthL; asthma; human; chromosome 11p; ASTHL1; ASTHL1; genetic locus;
KW   therapeutic; immunogen; ds.
XX
OS   Mus sp.
```

```
XX
FH   Key             Location/Qualifiers
FT   CDS             1..848
FT                   /*tag= a
FT                   /transl_except= (pos:847..848, aa:Gly)
FT                   /product= "asthlj protein"
FT                   /note= "the stop codon is not indicated"
XX
PN   WO9937809-A1.
XX
PD   29-JUL-1999.
XX
PF   21-JAN-1998; 98WO-US01260.
XX
PR   21-JAN-1998; 98WO-US01260.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Brooks-Wilson AR, Buckler A, Cardon L, Carey AH;
PI   Galvin M, Miller A, North M;
XX
DR   WPI; 1999-479058/40.
DR   P-PSDB; AAY41420.
XX
PT   Mammalian asthma related genes, useful for diagnosis of a
PT   predisposition to development of asthma
XX
PS   Disclosure; Page 186-187; 195pp; English.
XX
CC   The invention identifies a genetic locus ASTHL, associated with asthma,
CC   mapped to human chromosome 11p. ASTHL1 and ASTHL2 are genes present
CC   within the locus, located close to each other on human chromosome 11p,
CC   and have similar patterns of expression, and common sequence motifs. The
CC   ASTHL genes and fragments, encoded protein, genomic regulatory regions
CC   and anti-ASTHL antibodies are useful in the identification of individuals
CC   predisposed to development of asthma, and for the modulation of gene
CC   activity in vivo for prophylactic and therapeutic purposes. The ASTHL
CC   protein is useful as an immunogen to raise specific antibodies, in drug
CC   screening for compositions that mimic or modulate ASTHL activity or
CC   expression, including altered forms of ASTHL protein, and as a
CC   therapeutic. The present sequence represents a mouse asthlj
CC   cDNA sequence.
XX
SQ   Sequence 848 BP; 253 A; 226 C; 292 G; 167 T; 0 other;


alignment scores:
            Quality:  239.00              Length:     247
              Ratio:  1.897                 Gaps:       8
 Percent Similarity:  58.165   Percent Identity: 27.004

alignment block:
 US-09-126-945B-2 x AAZ18688   ..

 Align seq 1/1  to: AAZ18688  from: 1  to: 848

      147 ProMetAspTrpSerProSerAsnValGluLysTrpLeuLeuTrpThrGl 164
          |||:::  |||:::     :::|||  :::||||||
      139 CCTCAGTACTGGACCAAATACCAGGTGTGGGAATGGCTGCAGCACCTCCT 188

      165 uHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu.....L 178
                  :::|||                 |||||||||      :
      189 GGACACCAACCAGCTAGACGCTAGCTGCATCCCTTCCAGGAGTTCGACA 238

      179 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArg 194
          :::::|||::::::|||||||:::||||||  :::::::|||  :::
      239 TTAGCGGCAGAACACCTGTGCAGCATGAGTCTGCAGGAGTTCACGAGGGCA 288

      195 SerProLeuGlyGlyAspValLeuHisAlaHisLeu.............. 206
          :::      :::|||:::::::|||:::::::|||
      289 GCAGGCTCAGCTGGGCAGCTGCTCTACAGCAACCTACAGCATCTCAAGTC 338

      207 ..................AspIleTrpLysSerAla.........AlaT 214
                                 |||:::::::::|||||
      339 GAACGGCCAATGCAGCAGTGACCTTTTCCAGTCCGCACACAATGTCATTG 388

      214 rpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThr 230
                |||:::|||:::|||:::  ::::::  :::
      389 TCAAGACTGAACAAACCGATCCTTCCATCATGAAC.........ACATGG 429

      231 SerGluGluSerTrpThr..............AspSerGluValAsp.. 241
          :::||||||:::::::                 |||   ||||||
      430 AAAGAAGAAAACTATCTCTATGATCCCAGCTATGGTAGCACAGTAGATCT 479

      242 .............SerSerCysSerGlyGln..................... 247
                      :::   |||   :::|||
      480 GTTGGACAGTAAGACTTTCTGCCGGGCTCAGATCTCCATGACAACCTCCA 529

      247 ..................................................... 247

      530 GTCACCTTCCAGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCAC 579

      248 Profile.....................................HisLeuIr 252
          ||::::                                   |||||||
      580 CCTGTAAAGTCCCACACCAAAAAGCACAACCCAAGAGGCACTCACTTATG 629

      252 pGluPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPhet 269
          |:::|||::::::::::::|||||:::|||  :::  :::   :::|
      630 GCAGTTCATCCGAGACATTCTCTTGAGCCCAGACAAGAACCCAGGGCTGA 679

      269 leArgIrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAla 285
          |||:::|||  :::::::::::::|||||||||:::::      |||
      680 TCAAATGGCAAGACGGTTCGGAAGGCATCTTCAGGTTCCTCAAGTCAGAA 729

      286 GluValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTy 302
             |||||:::|||||||||   :::|||||||:::   :::|||:::||
      730 GCTGTGGCTCAGCTGTGGGGGAAAAAGAAAAATAACAGTAGCATGACATA 779

      302 rAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleA 319
          |:::|||||||||||||:::|||  |||||||||:::  |||:::
      780 CGAGAAGCTCAGCTGGGCTATGAGATATTACTACAAACGAGAAATCCTGG 829

      319 rgLysProAsp 322
          :::::  |||
      830 AACGTGTGGAT 840

seq_name: /C1=US/C2=ATTY/C3=09-126-945B/C4=2/C5=/C6=NA2000:ID:AAA80595

seq_documentation_block:
ID   AAA80595 standard; cDNA; 848 BP.
XX
AC   AAA80595;
XX
DT   22-NOV-2000 (first entry)
XX
DE   Mouse asthlj cDNA.
XX
KW   ASTHL locus; asthlj; mouse; murine; asthma;
KW   bronchial hyperreactivity; ots family; transcription factor;
KW   splice variant; genetic predisposition; polymorphism; antibody;
KW   drug screening; prophylaxis; therapy; diagnosis; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..848
FT                   /*tag= a
FT                   /partial
FT                   /product= "Mouse asthlj isoform alt"
FT                   /note= "No stop codon given in the specification"
FT                   /transl_except= (pos:847..848, aa:Arg)
XX
PN   US6087485-A.
XX
PD   11-JUL-2000.
XX
```

```
PF   21-JAN-1998;   98US-0009913.
XX
PR   21-JAN-1997;   97US-0035564.
PR   01-JUL-1997;   97US-0051432.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  North M,  Cardon L,  Buckler A;
PI   Brooks-Wilson AR,  Carey AH;
XX
DR   WPI; 2000-505109/45.
DR   P-PSDB; AAB11972.
XX
PT   New nucleic acids other than naturally occurring chromosomes encoding
PT   ASTHl protein, for e.g. screening compositions that modulate expression
PT   or function of ASTHl proteins or as diagnostics for genetic
PT   predisposition to asthma -
XX
PS   Disclosure; Column 251-254; 131pp; English.
XX
CC   The invention relates to the ASTHl locus on the short arm of human
CC   chromosome (11p). This locus comprises the ASTHll and ASTHlJ genes,
CC   which are associated with a genetic predisposition to asthma and
CC   bronchial hyperreactivity. The ASTHll and ASTHlJ genes are oriented in
CC   opposite directions with the ASTHl locus, and have similar patterns of
CC   expression and common sequence motifs. They are both expressed in
CC   trachea, lung and several other tissues. ASTHll and ASTHlJ are novel
CC   members of the ets family of transcription factors, which have been
CC   implicated in the activation of a variety of genes including the TCRa
CC   gene and cytokine genes known to be important in the aetiology of asthma.
CC   Both ASTHll and ASTHlJ mRNAs are alternatively spliced. Alternative
CC   splicing of transcripts has no effect on the open reading frame of
CC   ASTHlJ, as the exons involved are all 5' to the start [illegible]. In
CC   contrast, alternative splicing of ASTHll transcripts results in 8
CC   different ASTHll isoforms. The invention also encompasses mouse asthlj
CC   protein. The ASTHl nucleic acids are useful as diagnostics to identify a
CC   hereditary predisposition to asthma, as probes for identifying ASTHl
CC   related genes, for identifying expression of the gene in a biological
CC   specimen, and for generating genetically modified non-human animals or
CC   site specific gene modifications in cell lines. The encoded ASTHl
CC   proteins are useful as immunogens to raise specific antibodies, in drug
CC   screening for compositions that mimic or modulate activity or expression
CC   of ASTHll and/or ASTHlJ (ie. binding affinity, [illegible]), and
CC   as a therapeutic. The ASTHl genes or fragments thereof, encoded proteins,
CC   ASTHl genomic regulatory regions, and anti-ASTHll and anti-ASTHlJ
CC   antibodies are useful in the identification of individuals predisposed to
CC   development of asthma, and for modulation of gene activity in vivo for
CC   prophylactic and therapeutic purposes. The intact ASTHll or ASTHlJ
CC   proteins or active fragments thereof may be used to modulate or reduce
CC   bronchial hyperreactivity. The present sequence represents a murine
CC   asthlj cDNA.
XX
SQ   Sequence 848 BP; 258 A; 226 C; 262 G; 157 T; [illegible]

alignment_scores:
            Quality:  239.00              Length:    237
              Ratio:  1.897                 Gaps:      8
 Percent Similarity:  58.165   Percent Identity: 27.004

alignment_block:
 US-09-126-945B-2 x AAA80595   ..

 Align seq 1/2  to: AAA80595  from: 1  to: 848

      147 ProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGl 163
          ||:::   |||:::      :::|||   ::||||||
      139 CCTCAGTACTGGACCAAATACCAGGTGTGGGAATGGCTGCAGCACCTGCT 188

      163 uHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGln.....L 178
                  :::|||                  |||||||||      :
      189 GGACACCAACCAGCTAGACGCTAGCTGCATCCCTTTCCAGGAGTTCGACA 238

      178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGluPheArgGlnArg 194
          :::::|||:::::|||||::|||||   :::::|||   :::
      239 TTAGCGGAGAACACCTGTGCAGCATGAGCTGCAGGAGTTCACGAGGGCA 288

      195 SerProLeuGlyGlyAspValLeuHisAlaHisLeu..... 206
          :::     :::|||:::::|||:::::::|||
      289 GCAGGCTCAGCTGGGCAGCTGCTCTACAGCAACCTACAGCATCTCAAGTG 338

      207 .....AspIleTrpLysSerAla.....AlaI 214
                          |||:::::::::::|||||
      339 GAACGGCCAATGCAGCAGTGACCTTTCCAGTCCGCACACAATGTCATTG 388

      214 rpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThr 230
                ||:::::|||:::|||:::   ::::::      :::
      389 TCAAGACTGAACAAACCGATCCTTCCATCATGAAC.....ACATGG 429

      231 SerGluGluSerTrpThr.....AspSerGluValAsp.. 241
          :::||||||:::::::                 |||   ||||||
      430 AAAGAAGAAAACTATCTCTATGATGCCAGCTATGGTAGCACAGTAGATCT 479

      242 .....SerSerCysSerGlyGln..... 247
                     :::   |||   :::|||
      480 GTTGGACAGTAAGACTTTCTGCCGGGCTCAGATCTCCATGACAACCTCCA 529

      247 ..... 247

      530 GTCACCTTCCAGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCAC 579

      248 ProIle.....HisLeuTr 262
          |||:::                                  ||||||||
      580 CCTGTAAAGTCCCACACCAAAAAGCACAACCCAAGAGGCACTCACTTATG 629

      263 pGluPheLeuLysGluLeuGlnLeuLysProHisSerTyrGlyArgPhel 269
          |:::|||:::::::::::::|||||||:::|||   :::      :::|
      630 GGAGTTCATCCGAGACATTCTCTTGAGCCCAGACAAGAACCCAGGGCTGA 679

      269 leArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAla 285
          |||:::|||   :::::::::::::::|||||||||:::::::      |||
      680 TCAAATGGGAAGACCGTTCGGAAGGCATCTTCAGGTTCCTGAACTCAGAA 729

      286 GluValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnly 302
          ||||||:::|||||||||   :::|||||||:::   :::|||:::||
      730 G***T***T**A**TGT**G**AAAAA**AAAAA**AA**A**AC**A**A**A 779

      302 rAspLysLeuSerArgSerIleArgCysTyrTyrLysLysGlyIleIleA 319
          |:::|||||||||||:::::|||   |||||||||:::   ||:::::
      780 GGAGAAGCTCAGCCGGGCTATGAGATATTACTACAAACGAGAAATCCTGG 829

      319 rgLysProAsp 322
          :::::   |||
      830 AACGTGTGGAT 840

>>1  title  Sulfur transferase protein. [Homo sapiens]
     gb|AAV38454.1|AAV38454 DAT-AAV38454

seq_documentation_block:
ID   AAV38454 standard; cDNA; 1116 BP.
XX
AC   AAV38454;
XX
DT   09-NOV-1998  (first entry)
XX
DE   Transcription factor ESX gene open reading frame.
XX
KW   ESX, epithelial restricted with serine box, transcription factor;
KW   ETS; human; epithelial cancer, breast cancer, adenocarcinoma;
KW   diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            1..163
FT                   /*tag   a
FT                   /number   2
```

```
FT   exon            164..385
FT                   /*tag   b
FT                   /number   3
FT   exon            386..478
FT                   /*tag   c
FT                   /number   4
FT   exon            479..688
FT                   /*tag   d
FT                   /number   5
FT   exon            689..805
FT                   /*tag   e
FT                   /number   6
FT   exon            806..1001
FT                   /*tag   f
FT                   /number   8
FT   exon            1002..1111
FT                   /*tag   q
FT                   /number   9
FT   misc feature    1..189
FT                   /*tag   h
FT                   /product  first variable region
FT                   /note  "Claim 28"
FT   misc feature    190..309
FT                   /*tag   i
FT                   /product  pointed region
FT   misc feature    310..561
FT                   /*tag   j
FT                   /product  second variable region
FT                   /note  "Claim 8"
FT   misc feature    562..714
FT                   /*tag   k
FT                   /product  serine rich region
FT   misc feature    715..819
FT                   /*tag   l
FT                   /product  third variable region
FT                   /note  "Claim 28"
FT   misc feature    820..1062
FT                   /*tag   m
FT                   /product  Ets DNA binding domain
FT   misc feature    1064..1113
FT                   /*tag   n
FT                   /product  fourth variable region
FT                   /note  "Claim 28"
XX
PN   WO9824782 A2.
XX
PD   04 JUN 1998.
XX
PF   26 NOV 1997;   97WO-US21865.
XX
PR   25 NOV 1997;   97US-0031504.
PR   27 NOV 1996;   96US-0031504.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Benz CC;  Chang C;  Scott GK;
XX
DR   WPI; 1998-322755/28.
DR   P-PSDB; AAW60677.
XX
PT   Nucleic acid encoding the ESX transcription activator
PT   over expressed in epithelial, especially breast, cancers, also
PT   related polypeptide(s), antibodies, vectors and transformed cells,
PT   useful for diagnosis and treatment of cancer
XX
PS   Claim 8; Fig 1; 120pp; English.
XX
CC   This is the open reading frame of the newly isolated EIX gene that
CC   codes for a novel transcription factor (see AAW60677) associated
CC   with the aetiology of cancers, including epithelial cancers.
CC   The gene is a member of the ETS family and is designated EIX for
CC   epithelial restricted with serine box.  The gene is located at
CC   chromosome 1q32, a region amplified in 50% of early breast cancers.
CC   ESX is an early indicator of breast cancer, being expressed at the
CC   ductal carcinoma in situ stage.  Dysfunction of the ESX gene that
CC   results in increased expression is indicative of epithelial cancer
CC   (especially breast cancer but also gastric, ovarian and lung
CC   adenocarcinoma) and of unfavourable prognosis.  The ESX open
CC   reading frame was deduced from a cDNA clone of foetal liver spleen
CC   origin that was identified in an expressed sequence tag database
CC   search using human Ets 2, and from a clone obtained by 5'RACE of
CC   placental polyA mRNA.  Antibodies, antisense nucleic acids or
CC   inactive ESX mutant proteins that inhibit activity of ESX can be
CC   used to inhibit growth and proliferation of neoplastic cells,
CC   particularly cancer.  ESX polypeptides, polynucleotides and
CC   antibodies are also used to detect ESX, particularly for diagnosis
CC   and monitoring of cancer.  Polynucleotides may also be used to
CC   detect gross abnormalities in chromosome 1, e.g. for in utero
CC   screening for predisposition to cancer, and to generate ESX protein
CC   in vivo from gene therapy vectors.  ESX protein, genes and cDNA
CC   can also be used to screen for specific modulators, especially
CC   potential anticancer agents.
XX
SQ   Sequence 1116 BP; 259 A; 422 C; 435 G; 204 T; 0 other;


alignment scores:
            Quality:  238.50              Length:     888
              Ratio:  1.411                 Gaps:      15
 Percent Similarity:  43.552   Percent Identity: 22.680

alignment block:
 US-09-126-945B-2 x AAV38454   ..

 Align seq 1/1  to: AAV38454  from: 1  to: 1116

      55 LeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSe 71
         :::|||   ::::::   :::||||||:::   :::|||   ::::::::
      19 ATTAGCAACATTTTTAGCAACTACTTCAGTGCGATGTACAGCTCGGAGGA 68

      71 rSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluProProG 88
         :|||      |||:::::::|||   |||
      69 CTCCACCCTGGCCTCTGTTCCCCCTGCT..................... 96

      88 lnGluProGluGlnCysProValIleAspSerGlnAlaProAlaGlySer 104
                                                |||      ||||::
      97 ..................................GCCACCTTTGGGGCC 111

     105 LeuAspLeuValProGlyGlyLeuThrLeuGlu......GluHisSerLe 119
            ||||||||||          ||||||||||:::      :::   |||||
     112 GATGACTTGGTA.........CTGACCCTGAGCAACCCCCAGATGTCATT 152

     119 nGlnGlnValGln...SerMetValValGlyGlnValLeuLysAspIleG 135
         ||||   ::::::   :::      :::|||||||
     153 GGAGGCTACAGACAAGGCCAGCTGGTTGGGGGAA.............. 186

     135 lnThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSer 151
                                          :::|||:::   ||||||
     187 ..............................CAGCCCCAGTTCTGGTCC 204

     152 ProSerAsnValGlnLysTrpLeuLeuTrp......IhrGluHisGlnly 166
            :::::::|||      |||:::   :::          ::::::::||
     205 AAGACGCAGGTTCTGGACTGGATCAGTTACCAAGTGGAGAAGAACAAGTA 254

     166 rArgLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGlnL 183
         |          :::     :::      ::::::   |||      |
     255 CGACGCAAGCGCCATTGACTTCTCACGATCTGACATGGATGGCGCCACCC 304

     184 cnCysAlaMetSerGluGlnGlnPheArg...GlnArgSerProLeuGly 198
         |||||    :::   |||:::::::|||          :::|||||
     305 TCTGCAATTCTGCCCTTGAGGAGCTGCGTCTGCTCTTGGGCCTCTG... 351

     199 GlyAspValLeuHisAlaHisLeu...AspIleTrpLysSerAla..... 212
         |||||   |||||||||||:::|||   |||:::   :::|||:::
     352 GGGGACCAACTCCATGCCCAGCTGCGAGACCTCACTTCCAGCTCTTCTGA 401
```

```
     213 .......AlaTrpMet.....................................L 216
                 :::|||:::                                    :
     402 TGAGCTCAGTTGGATCATTGAGCTGCTGGAGAAGGATGGCATGGCCTTCC 451

     217 ysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGlu 232
         :|||      :::||||||   :::      :::      ::::::
     452 AGGAGGCCTAGACCCAGGGCCCTTTGACCAGGGCAGCCCTTTGCCCAG 501

     233 GluSerTrpThrAspSerGluValAspSer..............SerCy 244
         |||         |||:::::      |||                |||||
     502 GAGCTGCTGGACGACGGTCAGCAAGCCAGCCCCTACCACCCCGGGCAGCTG 551

     244 sSer..................................................... 245
         ::::
     552 TGGCGCAGCAGCCCCTCCCCTGGCAGCTGTGACGTCTCCACCCCAGGGA 601

     245 ..................................................... 245

     602 CTGGTGCTTCTCGGAGCTCCCACTCCTCAGACTCCGGTGGAAGTGACCTG 651

     245 ..................................................... 245

     652 GACCTGGATCCCACTGATGGCAAGCTCTTCGCCAGCGATGGTTTTCGTGA 701

     246 .....................................GlyGlnPro. 248
                                              ||||:|||
     702 CTGCAACAAGGGGGATCCCAAGCACGGGAAGCGGAAACGAGGCCGGCCC 751

     248 ..................................................... 248

     752 CAAAGCTGAGCAAAGAGTACTGGCACTGTCTCGAGGGCAAGAAGAGCAAG 801

     249 ..............IleHisLeuTrpGlnPheLeuLysGluLeuLeLe 260
                      |||||||||:::||:::::::::::|||::
     802 CACGCGCCCAGAGGCACCCACCTGTGGGAGTTCATCGGGCACATCCTCAT 851

     260 qLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysC 277
         :  ||:::      :::::::::|||   |||:::::::::|
     852 CCACCCGGAGCTCAACGAGGGCCTCATCAAGTGGGAGAATCGGCATGAAG 901

     277 lyIletheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIle 293
         |||:::|||||:::      |||      ||||||:::|||||||||
     902 GGGTCTTCAAGTTCCTGCGCTCCGAGGCTGTGGCCCAACTATGGGGCAA 951

     294 ArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleAr 311
         :::|||:::::      |||:::|||:::||||||||||||:::::||
     952 AAGAAAAAGAACAGCAACATGACCTACGAGAAGCTCAGCCGGGCCATCAG 1001

     310 qGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgL 327
         |   ||||||||||:::   |||:::::::::   |||   ::::::|||
    1002 GTACTACTACAAACGGGAGATCCTGGAACGGGTGGAT...GGCCGGCGAC 1048

     327 euValTyrGlnPhe 331
         :|||||:|:::|||
    1049 TCGTCTACAAGTTT 1062

seq_name: /SIS/projects/patseq/patseq/NA2000_DAT/AAZ49125

seq_documentation_block:
ID   AAZ49125 standard; DNA; 1116 BP.
XX
AC   AAZ49125;
XX
DT   06-APR-2000 (first entry)
XX
DE   Human ESX coding sequence.
XX
KW   ESX; epithelial-restricted with serine box, modulator screening, cancer,
KW   dysregulation detection; transcription; neoplastic cell growth inhibitor;
KW   diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9965929-A1.
XX
PD   23-DEC-1999.
XX
PF   15-JUN-1999;   99WO-US13277.
XX
PR   16-JUN-1998;   98US-0089409.
PR   09-APR-1999;   99US-0289535.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Benz CC, Scott GK, Chang C, Chao Y;
XX
DR   WPI; 2000-106080/09.
DR   P-PSDB; AAY69914.
XX
PT   Use of ESX (epithelial-restricted with serine box) for developing
PT   agents for diagnosis and treatment of disorders involving abnormal ESX
PT   regulation, particularly cancers
XX
PS   Example 1; Fig 1; 144pp; English.
XX
CC   This sequence encodes the human ESX (epithelial-restricted with serine
CC   box) protein. The invention relates to a method of screening for a
CC   modulator of ESX activity, comprising: (i) providing a target selected
CC   from a nucleic acid encoding a polypeptide of ESX exon 4 or ESX exon 7, a
CC   polypeptide comprising ESX exon 4 or exon 7; (ii) contacting the nucleic
CC   acid or polypeptide with a test agent; and (iii) detecting the binding of
CC   the test agent to the nucleic acid or polypeptide, where a test agent
CC   that binds to the sequence modulates ESX activity. The products and
CC   methods can be used for detecting dysregulation of an ESX gene as
CC   indicative of cancer, and for prognosis and selecting an appropriate
CC   treatment regime. They can also be used for activating or depressing
CC   transcription of a gene or cDNA and to identify nucleic acids and/or
CC   polypeptides that interact with ESX. Agents which inhibit the activity of
CC   ESX exon 4 or exon 7 can be used to inhibit the growth or proliferation
CC   of neoplastic cells, particularly cancer.
XX
SQ   Sequence 1116 BP; 255 A; 322 C; 345 G; 204 T; 0 other;


alignment_scores:
            Quality:  238.50              Length:     388
              Ratio:  1.411                 Gaps:      15
 Percent Similarity:  43.557   Percent Identity: 22.680

alignment_block:
 US-09-126-945B-2 x AAZ49125   ..

 Align seq 1/1  to: AAZ49125  from: 1  to: 1116

      55 LeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGlaAspSe 71
         :::|||   ::::::   :::|||||||:::   :::|||   ::::::::
      19 ATTAGCAACATTTTAGCAACTACTTCAGTGCGATGTACAGCTCGGAGGA 68

      71 rSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluProProG 88
         :|||      |||::::::|||   |||
      69 CTCCACCCTGGCCTCTGTTCCCCCTGCT..................... 96

      88 luGlnProGluGlnCysProValIleAspSerGlnAlaProAlaGlySer 104
                                              |||      ||||::
      97 .................................GCCACCTTTGGGGGCT 111

     105 LeuAspLeuValProGlyGlyLeuThrLeuGlu.....GluHisSerLe 119
           |||||||||         |||||||||:::      :::   |||||
     112 GATGACTTGGTA.........CTGACCCTGAGCAACCCCCAGATGTCATT 152

     119 uGluGlnValGln...SerMetValValGlyGluValLeuLysAspIleG 135
         ||||   ::::::   :::      :::|||||
     153 GGAGGGTACAGAGAAGGCCAGCTGGTTGGGGGAA.............. 186

     135 luThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSer 151
                                          :::|||:::   ||||||
```

```
187 ..............................CAGCCCCAGTTCTGGTCG 204

152 ProSerAsnValGluLysTrpLeuLeuTrp......ThrGluHisGlnIy 166
    :::::|||      |||:::   :::          ::::::::::||
205 AAGACGCAGGTTCTGGACTGGATCAGCTACCAAGTGCAGAAGAACAAGTA 254

166 rArgLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluL 183
    |          :::      :::      ::::::   |||      |
255 CGACGCAAGCGCCATTGACTTCTCACGATGTGACATGGATGGGCGCCACCC 304

183 cuCysAlaMetSerGluGluGlnPheArq...GlnArqSerProLeuGly 198
    |||||      :::   |||::::::|||         :::|||||
305 TCTGCAATTGTGCCCTTGAGGAGCTGCGTCTGGTCTTTGGGGCCTCTG... 351

199 GlyAspValLeuHisAlaHisLeu...AspIleTrpLysSerAla      212
    ||||||   ||||||||||:::|||   |||:::   :::|||:::
352 GGGGACCAACTCCATGCCCAGCTGCGAGACCTCACTTCCAGCTCTTCTGA 401

213 .......AlaTrpMet.................................L 216
          :::|||:::                                  :
402 TGAGCTCAGTTGGATCATTGAGCTGCTGGAGAAGGATGGCATGGCCTTCC 451

216 ysGluArqThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGlu 232
    ::|||      :::||||||   :::         :::      ::::::
452 AGGAGGCCCTAGACCCAGGGCCCTTTGACCACGGCAGCCCCTTTGCCCAG 501

233 GluSerTrpThrAspSerGluValAspSer.............SerCy 244
    |||         |||::::::     |||                |||||
502 GAGCTGCTGGACGACGGTCAGCAAGCCAGCCCCTACCACCCCGGCAGCTG 551

244 sSer.................................................. 245
    |:::
552 TGGCGCAGGAGCCCCCTCCCCTGGCAGCTCTGACGTCTCCACCGCAGGGA 601

245 .................................................... 245

602 CTGGTGCTTCTCGGAGCTCCCACTCCTCAGACTCCGGTGGAAGTGACGTG 651

245 .................................................... 245

652 GACCTGGATCCCACTGATGGCAAGCTCTTCCCCAGCGATGCTTTTCGTGA 701

246 ..................................GlyGlnPro. 248
                                        |||::::|||
702 CTGCAAGAAGGGGGATCCCAAGCACGGGAAGCGGAAACGAGGCCGGCCCC 751

248 .................................................... 248

752 GAAAGCTGAGCAAAGAGTACTGGGACTGTCTCGAGGGCAAGAAGAGCAAG 801

249 ..............IleHisLeuTrpGlnPheLeuLysGluLeuLeuLe 268
                 ||||||||||:::|||:::::::::::::|||:
802 CACGGCCCAGAGGCACCCACCTGTGGGAGTTCATCCGGGACATCCTCA1 851

269 uLysProHisSerTyrGlyArqPheIleArqTrpLeuAsnLysGluLysG 277
    :   |||:::          :::::::::|||   |||:::::::::|
852 CCACCCGGAGCTCAACGAGGGCCTCATGAAGTGGGAGAATCGGCATGAAG 901

277 lyIlePheLysIleGluAspSerAlaGlnValAlaArqLeuTrpGlyIle 293
    |:::||||||:::      |||      ||||||:::||||||||
902 GCGTCTTCAAGATCCTGGCGTCCGAGGCTGTGGCCCAACTATGGGGCCAA 951

294 ArqLysAsnArqProAlaMetAsnTyrAspLysLeuSerArqSerIleAr 310
    :::|||:::::::      |||:::|||:::|||||||||||||:::::||
952 AAGAAAAAGAACAGCAACATGAACTACGACAAGCTGAGCCGGGCCATCAG 1001

310 qGlnTyrTyrLysLysGlyIleIleArqLysProAspIleSerGluArqL 327
    |   ||||||||||:::   |||:::::::::   |||   :::::|||||
1002 GTACTACTACAAACGGCAGATCCTGCAACGGCTGGAT...GGCCGGCGAC 1048

327 euValTyrGlnPhe 341
    |||||||:::|||
1049 TCGTCTACAAGTTT 1062
```

```
seq_name: [???:??:?] aar?_gene=[?]?e=?=?=?p_NA2000:DAI:AAF15985

seq_documentation_block:
ID   AAF15985 standard; cDNA; 1884 BP.
XX
AC   AAF15985;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human prostate cancer antigen nucleotide sequence SEQ ID NO:420.
XX
KW   Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW   neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW   vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW   antibacterial; gene therapy; neural; immune; reproductive; renal;
KW   gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW   wound; infectious disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055174-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05988.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587513/55.
DR   P-PSDB; AAB56782.
XX
PT   Prostate cancer associated gene sequences, referred to as prostate
PT   cancer antigens, useful for treatment, prevention, and diagnosis of
PT   disorders such as prostate cancer -
XX
PS   Claim 1; Page 904; 2348pp; English.
XX
CC   AAF15566 to AAF16505 encode the human prostate cancer associated
CC   proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC   The prostate cancer antigens can have neuroprotective, cytostatic,
CC   cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC   nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC   and can be used in gene therapy. The prostate cancer antigen
CC   polynucleotides may be used for detection of prostate cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The prostate cancer antigens may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC   AAB57303 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1884 BP; 497 A; 488 C; 479 G; 416 T; 4 other;


alignment scores:
            Quality:  238.50              Length:    402
              Ratio:  1.491                 Gaps:     12
 Percent Similarity:  89.801   Percent Identity: 21.642

alignment_block:
 US-09-126-945B-2 x AAF15985   ..

 Align seq 1/1  to: AAF15985  from: 1  to: 1884

     100 AlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGlnGl 116
         ::|||:::  :::      :::  |||  |||:::  |||  ::
```

```
 87 FCACCAACTTGGAAGTCCINAATGTCCCCATGGGGGGTGTTCCTTCCAGA 86

116 uHisSer.....LeuGluGlnValGlnSerMetValV 127
    :  :::          |||          ::::::   ||||
 87 CTCCGGCAACTGTGAATTGCCTTTGTTAACCCCGTGCAGCAAGGCTGTGA 136

127 alGlyGluValLeuLysAspIleGluThrAlaCysLys..... 149
    ::::::::::::|||||          ::::::   |||
137 TGAGTCAAGCTTAAAAGCTACTTCAGTGGCTTCAAAAAGGAACAGCGG 186

140 LeuLeuAsuIleThrAlaAspProMetAspTrpSerProSerAsnValGl 156
       |||:::|||      :::|||      ||||||   ::::::|||
187 CGCCTGGGCATTCCAAACAACCCTGCCTGTGGAGTCAGCAACAGGTATG 236

156 nLysTrpLeuLeuTrpIhrGluHisGlnTyrArqLeuProProMetGlyL 175
    :::|||||||||||||:::   :::::::::   |||      :::   :
237 CCAGTGGCTTCTCTGGGCCACCAATGACTTCAGTCTGGTGAACGTGNAAI 286

175 ysAlaPheGlnGlu.LeuAlaGlyLysGluCysAlaMetSerGluGl 189
    ::|||      ::: :::   ||::::   ||||||   ::::::::::|
287 CTGCAGAGCTTCGGCATGAATGGCCAGATGCTGTGTAACCTTGGCAAGGA 336

189 uGlnPheArqGlnArqSerPro...LeuGlyGlyAspValLeuHisAlaH 205
    |::::::   :::   :::|||   :::   |||||::::||      |
337 ACGCTTCTGGAGCTGGCACCTGACTTGTGGGTGACATTCTCTGGGAAC 386

205 isLeuAsp..... 207
    |||||:::
387 ATCTGGACCAAATGATCAAAGAAAACCAAGAAAAGACAGAAGATCAATA1 436

208 .....IleTrpLysSerAlaAlaTrpMet..... 215
         :::  :::          ||::::
437 GAAGAAAATTCACACCTCAGCTCCGTTCCTCATTGGATTAACAGCAATAC 486

215 ..... 215

487 ATTAGGTTTTGGCACAGAGCAGGCGCCCTATGGAATGCAGACACAGAATT 536

215 ..... 215

537 ACCCCAAAGGCGGCCTCCTGGACAGCATGTCCGGCCTCCACACCCAGC 586

216 .....            IysGluArqIhrSe 230
                     ||::::|||   ||
587 GTACTCAGCTCTGAGCAGGAGTTCAGATGTTCCCCAAGTCTCGGCTCAG 636

220 rProGlyAlaIleHisTyrCysAlaSerThrSerGlu..... 242
    |    :::::   |||||:::   :::::::::
637 CTCCGTCAGCGTCACCTACTGCTCTGTCAGTCAGGACTTCCCAGGCAGCA 686

232 ..... 232

687 ACTTGAATTTGCTCACCAACAATTCTGGGACGCCCAAAGACCACGACTCC 736

233 .....GluSe 234
                     :::||
737 CCTGAGAACGGTGCGGCACAGCTTCGAGAGCTCAGACTCCCTGCTCCACTC 786

234 rTrpThrAsp.....S 248
    |||:::::::      :
787 CTGGAACAGCCAGTCGTCCTTGCTGGATGTGCAACGGGTTCCTTCCTCTG 836

238 erGluValAspSerSerCysSer..... 245
    :::::   :::::::::||||||
837 AGAGCTTCGAAGATGACTGCAGCCAGTCTCTCTGCCTCAATAAGCCAACC 886

245 ..... 245

887 ATGTCTTCAAGGATTACATCCAAGAGAGGAGTGACCCGGTGGAGCAAGG 936

246 .....GlyGlnP 248
                     :::  |
937 CAAACCAGTTATACCTGCAGCTGTGCTGGCCGGCTTCACAGGAAGTGGAC 986
```

```
248 roIleHisLeuTrpGlnPheLeuLysGluLeuLeuLysProHisSer 264
    |||||:::|||||||||||||||||   |||   |||||||:::      |||
987 CTATTCAGCTGTGGCAGTTTCTCCTGGAG...CTGCTATCAGACAAATCC 1033

265 TyrGlyArqPhoIleArqIrpLeuAsnLysGluLysGlyIleProLysIl 281
          ||||||   |||   :::   :::          |||||||::
1034 TGCCAGTCATTCATCAGCTGGACTGGA...GACGGATGGGAGTTTAAGCI 1080

281 eGluAspSerAlaGlnValAlaArqLeuTrpGlyIleArqLysAsnArqP 298
    :   |||      :::||||||||||   ||:|||   ||||||||||:::|
1081 CGCCGACGCCCGATCAGGTGGCCCGCCGTGGGGAAAGAGGAAAAATAAGC 1130

298 roAlaMetAsnTyrAspLysLeuSerArqSerIleArqGlnTyrTyrLys 314
    ||   ||||||||||:::||||||||||||||:::::|||   |||||
1131 CCAAGATGAACTACGACAAGCTGAGCCGGGGCTTACGCTACTATTACGAC 1180

315 LysGlyIleIleArqLysProAspIleSerGlnArqLeuValTyrGlnPh 331
    |||:::|||:|||:::|||   :::   :::::|||   ||||||:::||
1181 AAGAACATCATCCACAAGACGTCG...GGGAAGCGCTACGTGTACCGCTT 1227

331 eVal 332
    |||
1228 CGTG 1231

seq name: "5211..": rig:"...." q"...." g_NA"..." IAC AAQ91769

seq_documentation_block:
ID   AAQ91769 standard; cDNA to mRNA; 1800 BP.
XX
AC   AAQ91769;
XX
DT   09-FEB-1996  (first entry)
XX
DE   Coding sequence of PEA3 beta, an ETS transcription factor.
XX
KW   transcription factor; probe; reverse transcription; PCR; primer;
KW   expression vector; E.coli; COS cell; ras; cancer cell multiplication;
KW   polyoma virus; transformation; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1563
FT                   /*tag= a
FT                   /product= "PEA3 beta transcription factor"
FT   CDS             250..252
FT                   /*tag= b
FT                   /transl_except= (seq:CTC,aa:Pro)
FT   CDS             376..378
FT                   /*tag= c
FT                   /transl_except= (seq:AAC,aa:Lys)
FT   CDS             1498..1500
FT                   /*tag= d
FT                   /transl_except= (seq:ACC,aa:Ser)
XX
PN   JP07145197-A.
XX
PD   06-JUN-1995.
XX
PF   25-NOV-1993;   93JP-0295393.
XX
PR   25-NOV-1993;   93JP-0295393.
XX
PA   (EISA ) EISAI CO LTD.
PA   (HIRA/) HIRANO I.
XX
DR   WPI; 1995-237197/31.
DR   P-PSDB; AAR78185.
XX
PT   ETS transcription factor activated by ras - may be used in the study
PT   of cancer cell proliferation and the proliferation of the polyoma
PT   virus
```

```
XX
PS   Claim 2; Page 7 9; 9pp; Japanese.
XX
CC   The nucleotide sequence of the novel ETS transcription factor family
CC   member   PEA3-beta.  The gene was isolated from a HepG2 cell line cDNA
CC   library.  The probe for the screening was prepared by reverse
CC   transcription on HepG2 mRNA followed by PCR using primers AAQ91770 1,
CC   to produce a probe of 170 200 bp.  The screening isolated the full
CC   length sequence of the transcription factor.  The gene was inserted into
CC   the expression vectors pBluescript KS and pcDV1 for expression of the
CC   protein in E.coli and COS7 cells, respectively.  The ETS transcription
CC   factor has specificity for and is activated by ras. It is useful as a
CC   reagent in studies for the elucidation of the mechanism of cancer cell
CC   multiplication or polyoma virus transformation of cells.
XX
SQ   Sequence 1800 BP; 437 A; 522 C; 435 G; 405 T; 1 other;


Alignment scores:
            Quality:  236.00              Length:     881
              Ratio:  1.418                 Gaps:      18
 Percent Similarity:  46.982   Percent Identity: 25.722

Alignment block:
 US 09 126 945B 2 x AAQ91769   ..

 Align seq 1/1  to: AAQ91769  from: 1  to: 1800

        5 SerProGlyLeuSerSerValSerProSerHis.....LeuLeuLe 18
          :::|||      :::   :::||||||:::|||        |||:::
      403 ACCCCTCCTACAACCCCCCTCTCACCCACCCATCAGAATCCCCTATTTCC 452

       18 uProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyA 35
          |||||||||||      |||:::   :::      |||||||||
      453 CCCACCTCAGGCAACTCTGCCCACCTCAGGGCATGCCCCTGCAGCTGGCC 502

       35 laValGlyLeuGluArgArgAspTrpSerProSerProProAlaThrPro 51
          |||      :::              :::|||:::|||   :::   |||
      503 CAGTT.....CAAGGTGTGGGCCCCGCCCCCGCCCCCATTCGCTTCCA 546

       52 GluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrPr 68
          |||   |||                                     ||
      547 GAGCCTGGA.....CC 557

       68 uGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluG 85
          ||||||||||||||:::                              :
      558 ACAGCAGCAAACATTTGCGGTC.....CCCC 584

       85 luProProGluGluProGluGlnCysProValIle.....AspSerGln 99
          ::|||||:::::|||   |||   |||   :::      :::::|||
      584 GACCACCACATCAGCCCCTGCAGATGCCAAAGATGATGCCTGAAAACCAG 633

      100 AlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGl 116
          |||:::                                    ||||:
      634 TATCCAICA.....GAACA 647

      116 uHisSerLeuGluGlnVal.....Gln 124
          ::::      :::|||:::      |||
      648 GAGATTCAGAGACAACTGTCTGAACCCTGCCACCCCTTCCCTCCICAGC 697

      124 rMetValValGlyGluIleLeuLysAspIleGluThrAlaCysLysLeu 140
          |||   |||:::          :::   :::          :::
      698 CAGGAGTTCCTGGAGATAATCGCCCCAGTTACCATCGGCAAATGTCAGAA 747

      141 LeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLy 157
          :::   |||   |||            |||:::::   :::::
      748 CCTATTGTCCCTGCAGCTCCCCCG.....CCCCCTCAGGGATTCAAACA 791

      157 sTrpLeuLeuTrpThrGluHisGlnTyrArgLeuPro.....P 170
          :           :::::::::   :::|||           |
      792 AGAATACCATGACCCACTCTATGAACATGGGGTCCCGGGCATGCCAGGGC 841

      170 roMetGlyLysAlaPheGlnGluLeuAlaGly.....Lys 181
```

```
          ||   :::   :::||||||:::      |||           :::
      842 CCCCAGCACACAGGGTTCCAGTCACCAATGGGAATCAAGCAGGAGCCTGG 891

      182 GluLeuCysAlaMetSerGluGluGlnPheArgGluArgSerProLeuGl 198
          :::   |||:::   ||||||           |||   |||   :::
      892 GATTACTGCGTCGATTCAGAAGTGCCTAACTGCCAGTCATCCTACATGAG 941

      198 yGlyAspValLeuHisAlaHisLeuAspProTrpLysSerAlaAlaTrpM 215
          |||      :::
      942 AGGGGTTAITTC..................................... 954

      215 etLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSer 231
                                                |||:::|||
      955 .....TCCAGCAGC 963

      232 GluGluSerTrpThrAspSerGlu.....ValAspSerSe 243
          :::|||:::::::::   ::::::              |||::::::
      964 CAIGAAGCTTTTCATATGAAAAAGATCCCCGATTATACTTTGACGACAC 1013

      243 rCys..... 244
          :|||
     1014 TTGTGTTGTGCCTGAGAGACTGGAAGCCAAAGTCAAACAGGAGCCTACCA 1063

      245 .....SerGlyGluPro.....IleHisLeuTrp 252
               :::|||   |||           :::::||||||
     1064 TGTATCGAGAGGGGCCCCCTTACCAGAGGCGAGGTTCCCTTCAGCTGTGG 1113

      253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl 269
          |||||||||      ||||||      |||   :::   :::::||||||
     1114 CAGTTCCTGGTCAGCCTTCTTGATGACCCA...GCCAATGCCCACTTCAT 1160

      269 eArgTrpLeuAsnLysGluLysGlyIle...PheLysIleGluAspSerA 285
          |   |||   :::       :::|||:::   |||||||:::   :::
     1161 TGCCTGGACAGGT.....CGAGGCATGGAGTTCAAGCTGATAGAACCGG 1204

      285 laGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsn 301
          :::||||||||   ||||||||||:::|||||||||||||||||||||||
     1205 AAGAGGTTGCTCGGCGCTGCGGCATCCAGAAGAACCGGCCAGCCATGAAC 1254

      302 TyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIl 318
          ||||||||||||||||||||||:::|||   ||||||:::||||||||::
     1255 TATGACAAGCTGAGCGGCTCTATCCGCTATTACTATGAAAAGGGCATCAT 1304

      318 eArgLysProAspIleSerGluArgLeuValTyrGlnPheVal 332
          :::|||      :::::::||   ||||||:::||||||
     1305 GCAGAAGGTG...GCTGGAGACGGATACGTCTACAAATTTGTC 1344

Seq_name: /SEQ_Seqlist/pesseq/pem-seqs/NA/998.DAT/AAV61167

Seq_documentation block:
ID   AAV61167 standard; cDNA; 852 BP.
XX
AC   AAV61167;
XX
DT   06-JAN-1999  (first entry)
XX
DE   cDNA sequence of prostate tumour clone PIB.
XX
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
XX
OS   Homo sapiens.
XX
PN   WO9837093-A2.
XX
PD   27-AUG-1998.
XX
PF   25-FEB-1998;   98WO-US003492.
XX
PR   09-FEB-1998;   98US-0020956.
PR   25-FEB-1997;   97US-0806099.
PR   01-AUG-1997;   97US-0904804.
XX
PA   (CORI-) CORIXA CORP.
```

```
XX
PI   Dillon DC,  Xu J;
XX
DR   WPI; 1998-609886/51.
XX
PI   Polypeptides comprising immunogenic portions of prostate proteins -
PI   used in a vaccine for the treatment of prostate cancer
XX
PS   Claim 3; Page 53; 130pp; English.
XX
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
XX
SQ   Sequence 852 BP; 174 A; 195 C; 206 G; 274 T; 3 other;


alignment_scores.
            Quality:  244.50              Length:     212
              Ratio:  1.954                 Gaps:       8
 Percent Similarity:  56.604   Percent Identity: 27.830

alignment_block:
 US-09-126-945B-2 x AAV61167/rev   ..

 Align seq 1/1  to reverse of: AAV61167  from: 1  to: 852

      177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGl 193
          ::::::   ||||::::::|||||||::|||||||   ::::::|||   ::
      775 GACATCAACGGCGAGCACCTTTGCAGCATGAGTTTGCAGGAGTTCACCCG 726

      193 nArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu..... 206
          :   :::      :::|||   :::|||::::::::::|||
      725 GGCGGCAGGGACGGCGGGGCANCTCCTCTACAGCAACTTGCAGCATCTGA 676

      207 .....AspIleTrpLysSer.....Ala 212
                          |||::::::::::|||         :::
      675 AGTGGAACGGCCAGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGTC 626

      213 AlaTrpMetLysGluArgThrSerProGlyAlaIle..... 224
                |||::::|||::::|||   :::
      625 ATTGTCAACAGTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAGA 576

      225 ....HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluV 240
              ::|||      :::|||:::      |||   ::|||      :
      575 CNAGAACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG...T 529

      240 alAspSerSer.....CysSerGlyGln..... 247
          ::|||||:::      |||   ::|||
      528 TGGACAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAGT 479

      248 ......ProIle..... 249
                .||:::
      478 CACCTTCCTGTTGCAGATCACCTGATATCAAAAACCACCAACACCCCCC 429

      250 .....HisLeuTrpG 258
                        |||||||||:
      428 TGCCAAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGG 379

      259 laPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 269
          ::|||||||||||||||||||||||   :::          :::||||
      378 AATTCATCGGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATA 329

      270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 286
          ::::||   :::::::::::::|||:::|||:::::      |||
      328 AAAATGGGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGC 279

      286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 303
           ||||||:::||||||||   :::|||||||:::   :::||||:::|||
      278 AGTGGCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTATG 229

      303 splysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleIleArg 319
          ::||||||||||||||:::::|||   ||||||||||:::   |||::::::
      228 AAAAGCTCAGCTGAGCTATGAGATATTACTACAAAAGAGAAATTCTGGAG 179

      320 LysProAspIleSerGluArgLeuValTyrGlnPhe 331
          :::   |||   :::::||||||||||||||::|||
      178 CGTGTGGAT...GGACGAAGACTGGTATATAAATTT 146

seq_name: US-09-126-945B-2 x AAV58521/rev  pg_NA1999 DAT-AAV58521

seq_documentation_block:
ID   AAV58521 standard; cDNA; 852 BP.
XX
AC   AAV58521;
XX
DT   08-DEC-1998  (first entry)
XX
DE   Prostate tumour specific gene clone P18.
XX
KW   Prostate tumour specific gene; human; prostate cancer, detection;
KW   therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9847418-A2.
XX
PD   27-AUG-1998.
XX
PF   25-FEB-1998;   98WO-US003690.
XX
PR   09-FEB-1998;   98US-0090480P.
PR   25-FEB-1997;   97US-0806596.
PR   01-AUG-1997;   97US-0904809.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillon DC,  Xu J;
XX
DR   WPI; 1998-480805/41.
XX
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
XX
PS   Claim 1; Page 55; 141pp; English.
XX
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
XX
SQ   Sequence 852 BP; 174 A; 195 C; 206 G; 274 T; 3 other;


alignment_scores:
            Quality:  244.50              Length:     212
              Ratio:  1.954                 Gaps:       8
 Percent Similarity:  56.604   Percent Identity: 27.830

alignment_block:
 US-09-126-945B-2 x AAV58521/rev   ..

 Align seq 1/1  to reverse of: AAV58521  from: 1  to: 852

      177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGl 193
          ::::::   ||||::::::|||||||::|||||||   ::::::|||   ::
      775 GACATCAACGGCGAGCACCTTTGCAGCATGAGTTTGCAGGAGTTCACCCG 726
```

```
193 nArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu..... 206
    :   :::      :::|||   :::|||||||||||||
725 GGCGGCAGGGACGGCGGGGCANCTCCTCTACAGCAACTTGCAGCATCTGA 676

207 .....AspIleTrpLysSer.....Ala 212
                 |||:::::::::|||       :::
675 AGTGGAACGGCCACGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGTC 626

213 AlaTrpMetLysGluArgThrSerProGlyAlaIle..... 224
          |||:::|||::|||||:   :::
625 ATTGTCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAGA 576

225 ....HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluV 240
    :::|||     :::|||::     |||   :::|||     :
575 CNAGAACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG...T 529

240 alAspSerSer.....CysSerGlyGln..... 247
    ::|||||:::     |||   :::|||
528 TGGACAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAGT 479

248 .....ProIle..... 249
         |||:::
478 CACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCCC 429

250 .....HisLeuIrpG 254
                    |||||||||:
428 TGCCAAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGG 379

253 InPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 269
    ::|||:::::::::::::|||||:::|||   :::          :::|||
378 AATTCATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATA 329

270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 286
    :::|||   :::::::::::|||:::|||:::::      |||
428 AAATGGGAACACCCATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGC 379

286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 303
     |||||:::||||||||   :::|||||:::   :::|||:::|||
278 AGTGGCTCAGCTATGGGCTAAAAACAAGAACAACAGCAGCATGACCTATG 229

303 spLysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleIleArg 319
    ::|||||||||||||:::::|||   |||||||||:::   |||::::::
228 AAAAGCTCAGCCGAGCTATCAGATATTACTACAAAAGAGAAATTCTGGAG 179

320 LysProAspIleSerGlnArgLeuValTyrGlnPhe 341
    :::   |||   ::::::|||||||||||:::|||
178 CGTGTGGAT...GGACGAAGACTGGTATATAAATTT 146
```

```
seq name: _??E9-8?e pla?a?ese seq?pro sequ/NA2000.DAT:AAA06284

Seq documentation block:
ID   AAA06284 standard; cDNA; 852 BP.
XX
AC   AAA06284;
XX
DT   13-JUN-2000 (first entry)
XX
DE   Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:44.
XX
KW   Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW   immunogenic; cytostatic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200004149-A2.
XX
PD   27-JAN-2000.
XX
PF   14-JUL-1999;   99WO-US15848.
XX
PR   14-JUL-1998;   98US-0115453.
PR   14-JUL-1998;   98US-0116134.
PR   23-SEP-1998;   98US-0150812.
PR   23-SEP-1998;   98US-0159822.
PR   15-JAN-1999;   99US-0232149.
PR   15-JAN-1999;   99US-0232880.
PR   09-APR-1999;   99US-0288946.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillon DC;  Barlocker SL;  Yuqiu J;  Xu J;  Mitcham JL;
XX
DR   WPI; 2000-171268/15.
XX
PT   New polypeptide useful for treating and diagnosing prostate cancer
PT   comprises an immunogenic portion of prostate tumor protein
XX
PS   Claim 1; Page 114-115; 264pp; English.
XX
CC   The present invention describes isolated polypeptides, comprising an
CC   immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC   and polynucleotides encoding them have cytostatic activity and can be
CC   used in vaccines and in gene therapy. The polypeptides and
CC   polynucleotides encoding them, antigen presenting cells which express
CC   the polypeptides, antibodies against the polypeptides and vaccines
CC   comprising them can be used for inhibiting the development of prostate
CC   cancer in a patient. The polypeptides can be used to generate antibodies
CC   or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC   the polynucleotides encoding the polypeptides can be used as a probe or
CC   to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC   AAY82000 to AAY82020 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 852 BP; 174 A; 195 C; 206 G; 274 T; 3 other;


alignment_scores:
            Quality:  234.50              Length:     212
              Ratio:  1.954                 Gaps:       8
 Percent Similarity:  56.604   Percent Identity: 27.830

alignment_block:
 US-09-126-945b-2 x AAA06284/rev   ..

 Align seq 1/1  to reverse of: AAA06284  from: 1  to: 852

     177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGlnGlnGlnPheArgGl 193
         ::::::   |||:::::::||||||:::||||||   ::::::|||   ::
     775 GACATCAACGGCGAGCACTTTGCAGCATGAGTTGCAGGAGTTCACCCG 726

     193 nArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu..... 206
         :   :::      :::|||   :::|||||||||||:::
     725 GGCGGCAGGGACGGCGGGGCANCTCCTCTACAGCAACTTGCAGCATCTGA 676

     207 .....AspIleTrpLysSer.....Ala 212
                      |||:::::::::|||       :::
     675 AGTGGAACGGCCACGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGTC 626

     213 AlaTrpMetLysGluArgThrSerProGlyAlaIle..... 224
               |||:::|||::|||||:   :::
     625 ATTGTCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAGA 576

     225 ....HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluV 240
         :::|||     :::|||::     |||   :::|||     :
     575 CNAGAACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG...T 529

     240 alAspSerSer.....CysSerGlyGln..... 247
         ::|||||:::     |||   :::|||
     528 TGGACAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAGT 479

     248 .....ProIle..... 249
              |||:::
     478 CACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCCC 429

     250 .....HisLeuIrpG 254
                         |||||||||:
```

```
      428 TGCCAAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGG 479

      254 luPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 269
          :::|||||||||||||||||||||||:|||   :::          :::|||
      480 AATTCATCCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATA 529

      270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGlnAspSerAlaGl 289
          :::|||   ::::::::::::|||:::|||::::::      |||
      530 AAAIGGGAAGACCGATCTGAGGGCGICTTCAGGTTCTTGAAATCAGAGGC 579

      286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 335
            ||||||:::|||||||||   :::|||||:::   :::|||:::|||
      580 ACTGCCTCAGCTATGGGCTAAAAAGAAGAACAACAGCAGCATGACCTATC 629

      304 spLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 353
          :::|||||||||||||:::::::||   |||||||||:::   |||::::
      630 AAAAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATTCTGGAG 679

      320 LysProAspIleSerGlnArgLeuValTyrGlnPhe 351
          :::   |||   :::::|||||||||||||||||||
      680 CGTGTGGAT...GGACGAAGACTGGTATATAAATTT 716

seq_name=> ...5B_2_x_AAA61646_1_1740_1_NA_0_0_FAT_AAA61646

seq_documentation_block:
ID   AAA61646 standard; cDNA; 1740 BP.
XX
AC   AAA61646;
XX
DT   23-OCT-2000  (first entry)
XX
DE   cDNA encoding rat Elf-1 polymorphic variant #2.
XX
KW   Elf-1; transcriptional regulatory factor; rat; polymorphic variant;
KW   immunological disease; allergic disease; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1740
FT                   /*tag=  a
FT                   /product= "Rat Elf-1 variant #2"
FT                   /transl_except= (pos:397..399, aa:Asp)
FT                   /transl_except= (pos:445..447, aa:Glu)
XX
PN   JP2000135088 A.
XX
PD   16-MAY-2000.
XX
PF   30-OCT-1998;   98JP-0309595
XX
PR   30-OCT-1998;   98JP-0309595.
XX
PA   (ASAK ) ASAHI BREWERIES LTD.
XX
DR   WPI; 2000-425802/37.
DR   P-PSDB; AAB04191.
XX
PT   A transcription regulating factor Elf-1 cDNA and its polymorphism
XX
PS   Claim 5; Page 10-13; 21pp; Japanese.
XX
CC   The invention relates to three polymorphic variants of a rat
CC   transcriptional regulatory factor Elf-1 (AAB03190-B03192) and nucleic
CC   acids encoding them (AAA61645-A61647). The invention also encompasses
CC   the use of Elf-1 for regulating the transcription of a gene in an
CC   expression vector. The cDNA encoding rat Elf-1 variant #1 (AAA61645)
CC   was isolated using PCR primers (AAA61648-A61649) based on mouse Elf-1
CC   cDNA, and the rat Elf-1 variant #2 and #3 cDNAs (AAA61646-A61647) were
CC   subsequently isolated. The Elf-1 cDNA and protein sequences are
CC   potentially useful for the prevention and treatment of immunological and
CC   allergic diseases mediated by Elf-1. The present sequence represents
CC   cDNA encoding a variant (#2) of rat Elf-1.
XX
SQ   Sequence 1740 BP; 518 A; 445 C; 397 G; 380 T; 0 other;


alignment_scores:
            Quality:  233.50              Length:     86
              Ratio:  4.538                 Gaps:      1
 Percent Similarity:  76.744   Percent Identity: 50.000

alignment_block:
 US-09-126-945B-2 x AAA61646  ..

 Align seq 1/1  to: AAA61646  from: 1  to: 1740

      246 GlyGluProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 262
          |||:::  |||:::|||||||:::|||||      ||||||
      502 GGAAACACAATTTACCTTTGGGAGTTTTTACTGGCTCTGCTTCAGGACAA 551

      262 cHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePh 279
          :::      ::::::|||:::|||   :::::|||||||||||||
      552 AGCTACCTGTCCTAAGTACATCAAATGGACACAGCGAGAAAAAGGCATTT 601

      279 heLysIleGlnAspSerAlaValAlaArgLeuTrpGlyIleArgLys 295
          ||||:::   ||||||      |||:::|||||||||||||   :::|||
      602 TTAAGTTGGTCGACTCTAAAGCTGTGTCCAGATTATGGGGGAAGCACAAA 651

      296 AsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTy 412
          |||:::|||  ||||||||||:::  :::::|||:::::::|||  ||
      652 AACAAACCTGACATGAACTATGACAAGCTGAGCAGATCTATCAGGTACTA 701

      312 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVal 329
          ||||:::::|||||||:::  |||  :::  :::|||||||||||||
      702 TTACCAAAGAGGTATTCTTGCAAAAGTGGCAA...GGTCAGCGCTTGGTGT 748

      329 yrGlnPhe 336
          |||||||
      749 ATCAGTTT 756

seq_name=> ...5B_2_x_AAA61645_1_2744_1_NA_0_0_FAT_AAA61645

seq_documentation_block:
ID   AAA61645 standard; cDNA; 2744 BP.
XX
AC   AAA61645;
XX
DT   23-OCT-2000  (first entry)
XX
DE   cDNA encoding rat Elf-1 polymorphic variant #1.
XX
KW   Elf-1; transcriptional regulatory factor; rat; polymorphic variant;
KW   immunological disease; allergic disease; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..207
FT                   /*tag=  a
FT   CDS             208..2055
FT                   /*tag=  b
FT                   /product= "Rat Elf-1 variant #1"
FT   3'UTR           2056..2744
FT                   /*tag=  c
XX
PN   JP2000135088 A.
XX
PD   16-MAY-2000.
XX
PF   30-OCT-1998;   98JP-0309595.
XX
PR   30-OCT-1998;   98JP-0309595.
XX
PA   (ASAK ) ASAHI BREWERIES LTD.
XX
```

```
DR   WPI; 2000-425802/37.
DR   P-PSDB; AAB03190.
XX
PT   A transcription regulating factor Elf 1 cDNA and its polymorphism
XX
PS   Claim 4; Page 6-8; 21pp; Japanese.
XX
CC   The invention relates to three polymorphic variants of a rat
CC   transcriptional regulatory factor Elf 1 (AAB03190 B03192) and nucleic
CC   acids encoding them (AAA61645 A61647). The invention also encompasses
CC   the use of Elf 1 for regulating the transcription of a gene in an
CC   expression vector. The cDNA encoding rat Elf-1 variant #1 (AAA61645)
CC   was isolated using PCR primers (AAA61648 A61649) based on mouse Elf 1
CC   cDNA, and the rat Elf 1 variant #2 and 3# cDNAs (AAA61646-A61647) were
CC   subsequently isolated. The Elf 1 cDNA and protein sequences are
CC   potentially useful for the prevention and treatment of immunological and
CC   allergic diseases mediated by Elf 1. The present sequence represents
CC   cDNA encoding a variant (#1) of rat Elf 1.
XX
SQ   Sequence 2744 BP; 816 A; 653 C; 588 G; 687 T; 0 other;


alignment scores:
            Quality:  233.50              Length:      86
              Ratio:  3.538                 Gaps:       1
 Percent Similarity:  76.744   Percent Identity: 50.000

alignment block:
 US-09-126-945B-2 x AAA61645   ..

 Align seq 1/1  to: AAA61645  from: 1  to: 2744

      246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 292
          |||:::   |||:::||||||::||||||      ||||||
      817 GGAAACACAATTTACCTTTGGGAGTCTTTGCTGGCTCTGCTTCAGGACAA 866

      262 nHisSerTyrGlyArgPheIleArgProLeuAsnLysGluLysGlyIleP 279
              :::     ::::::|||:::|||   ::::::||||||||||||||
      867 AGCTACCTGTCCTTAAGTACATCAAATGGACACAGCGAGAAAAAGGCATTT 916

      279 heLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLys 296
          |||||||:   ||||||      |||:::||||||||||||   ::::||
      917 TTAAGTTGGTCGACTCTAAAGCTGTGTCCACATTATGGGGCAAGCACAAA 966

      296 AsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTy 312
          |||:::|||   ||||||||||:::   ::::::|||:::::|||   ||
      967 AACAAACCTGACATGAACTATGACAAGCTATCGGGCGAGCACTCAGGTACTA 1016

      312 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTr 329
          ||||:::::|||||:::   |||   :::   ::|||||||||||||||
     1017 TTACCAAAGAGGTATTCTTGCAAAACTGGAA...GGTCAGCGCTTGCTGTT 1063

      329 yrGlnPhe 331
          |||||||
     1064 ATCACTTT 1071

seq name: /cygdrive/p/data/pto/seq/pcorsepo/SAI/06/US-09-126-945B-2.AA.s7087

seq documentation block:
ID   AAI37087 standard; cDNA to mRNA; 2064 BP.
XX
AC   AAI37087;
XX
DT   25-APR-1997  (first entry)
XX
DE   ElAF matrix metalloproteinase regulator, cDNA.
XX
KW   ElAF; matrix; metalloproteinase; regulator; infiltration; cancer;
KW   metastasis; cell; control; antisense; decoy; DNA binding region;
KW   target DNA; ribosome, induction, diagnosis, detection; treatment;
KW   mammary cancer; fibrosarcoma; osteosarcoma; lung cancer; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1389
FT                   /*tag= a
XX
PN   WO9624679-A1.
XX
PD   15-AUG-1996.
XX
PF   09-JAN-1996;   96WO-JP00016.
XX
PR   09-FEB-1995;   95JP-0020173.
XX
PA   (IAKI ) TAKARA SHUZO CO LTD.
XX
PI   Fujimaga K,  Higashino F,  Yoshida K;
XX
DR   WPI; 1996-384227/38.
DR   P-PSDB; AAW00167.
XX
PT   Control of cancer cell infiltration by ElAF gene expression
PT   regulation - also diagnosis of cancer by detection of ElAF gene
PT   expression products
XX
PS   Example 5; Pages 38-42; 65pp; Japanese.
XX
CC   The present sequence encodes the ElAF protein, which is a matrix
CC   metalloproteinase regulator. The infiltration and metastasis of
CC   cancer cells can be controlled by regulating the expression and
CC   expression products of the ElAF gene. This may be accomplished by
CC   inducing antisense DNA or RNA for the ElAF gene, a decoy gene
CC   expressing the DNA binding region of the ElAF protein, the target
CC   DNA for the DNA binding region of the ElAF protein or ribosomes
CC   corresponding to the ElAF gene mRNA. Cancer can be diagnosed by
CC   detecting ElAF gene expression products, e.g. ElAF protein or mRNA.
CC   These methods may be used in the treatment and diagnosis of cancer,
CC   e.g. mammary cancer, fibrosarcoma, osteosarcoma, lung cancer, etc..
XX
SQ   Sequence 2064 BP; 427 A; 648 C; 569 G; 420 T; 0 other;


alignment scores:
            Quality:  230.50              Length:     357
              Ratio:  1.332                 Gaps:      21
 Percent Similarity:  48.459   Percent Identity: 28.571

alignment block:
 US-09-126-945B-2 x AAI37087   ..

 Align seq 1/1  to: AAI37087  from: 1  to: 2064

       44 SerProSerProProAlaThrProGluGluGlyLeuSerAlaPheTyrLe 60
          |||   :::||||||          :::::|||   ::::::::::|||
      265 AGCAGGAAGCCGCCACTCCCCTACCACCATGGCGACGCAGCCCTTACTC 314

       60 nSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysA 77
          |||   :::|||          |||:::::   :::      :::   |
      315 CAGTGCCTATCACCCC.....CCCAGACAAATCGCCATCAAGTCCCCTG 358

       77 laProGlyAla.....SerSerArgGluGluProProGluGluProGlu 91
          ||||||||||      ::::|||      ::::|||      :::   |||
      359 CCCCTGGTGCCCTTGGACAGTCGCCCCCTACAGCCCTTTCCCCGGGCAGAG 408

       92 GlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuVa 108
          |||         :::   |||         ::|||:::   :::
      409 CAACGGAATTTCCTGAGATCC.....TCTGGCACCTCCCAGCCCCA 449

      108 lProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerM 125
          ||||||         |||   ||||||||
      450 CCCTGGCCAGGGTACCTCGGGGAACATAGC..................... 480

      125 etValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeu 141
```

```
     480 ..................................................  480

     142 AsnIleThrAlaAspProMetAspTrpSerProSerAsnValGluLysPr  156
         ::::::      :::|||:::|||         |||
     481 TCCGTTTTCCAGCAGCCGCCTGGACATTTGCCACTCC..............  516

     158 pLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaP  175
             :::|||:::::                 |||:::       :
     517 .......TTCACATCTCAGGGAGGGGCCGGGAACCCCTGCCAGCCCCCT  559

     175 heGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPhe  191
         :::|||:::      ::::::                 ::::::::::|||
     560 ACCAACACCAGCTGTCGGAGCCCTGCCCACCCTATCCCCAGCAGAGCTTT  609

     192 ArgGln......ArgSerProLeu.....................  197
         :::|||      ::::::||||||
     610 AAGCAAGAATACCATGATCCCCTGTATGAACAGCGGGGCCAGCCAGCCCT  659

     198 .......GlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerA  212
             ||||||      ::::::::::|||         :::   :::|
     660 GGACCAGGGTGGG......GTCAATGGGCACAGG......TACCCAGGGG  697

     212 laAlaTrpMetLysGluArgThr.....................  219
         || :::|||   :::::::|||:::
     698 CGGGGGTGCTGATCAAACAGCAACAGACGGACTTCGCCTACGACTCAGAT  747

     220 .SerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSer.Trp  235
          ||||||||||||   |||   |||:::|||||||:::::   |||:|||
     748 GTCACCGGGTGCG...CATCAATGTACCTCCACACACAGGGCTTCTCTGG  794

     236 IhrAspSerGlu.....................................  239
         :::   |||:::
     795 GCCCTCTCCAGGTCACGGGGCCATGGGCTATGGCTATCAGAAACCTCTGC  844

     240 .......Val.AspSerSerCys.....................  244
                |||:::   |||
     845 GACCATTCCCAGATGATGTCTGCGTTGTCCCTGAGAAATTTGAAGGAGAC  894

     245 ..............................SerGlyGlnPro.....  219
                                      :::|||   |||
     895 ATCAAGCAGCAAGGGGTCGGTGCATTTCGAGAGGGCCGCCCTACCAGGG  944

     249 ..........IleHisLeuTrpGlnPheLeuLysGlnLeuLeuLeuLysP  262
                 :::::|||||||||||||||||      ||||||       |
     945 CCGGGGTGCCCTGCAGCTGTGGCAATTTCTGGTGGCCTGCTGGATGACC  994

     262 roHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle  279
         ||   :::   :::::::::|||   |||   :::       :::||:::
     995 CA...ACAAATGCCCATTTCATTGCCTGGACGGGC......CGGGGAATG 1035

     279 ...PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleAr  294
           |:|||:::   :::       :::|||||||||||||||||||||:::
    1036 GAGTTCAAGCTCATTGAGCCTGAGGAGGTCGCCAGGCTCTGGGGCATCCA 1085

     294 qLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgT  311
         :|||||||||||||||||||||||||||||||||||||||||||||:::|||
    1086 GAAGAACCGGCCAGCCATGAATTACGACAAGCTGAGCCGCTCGCTCCGAT 1135

     311 InTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeu  327
           |||||:::|||||||||||:::::|||         :::::|||
    1136 ACTATTATGAGAAAGGCATCATCCACAAGGTG...GCTGGTGAGCGTTAC 1182

     328 ValTyrGlnPheVal 332
         ||||||::||||||
    1183 GTGTACAAGTTTGTC 1197

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ50016

seq_documentation_block:
ID   AAZ50016 standard; cDNA; 2064 BP.
XX
AC   AAZ50016;
XX
DT   25-APR-2000  (first entry)
XX
DE   Human polyomavirus enhancer activator 3 cDNA.
XX
KW   Human polyomavirus enhancer activator 3; PEA3; transformation;
KW   tumorigenic; metastatic; cancer; neu-mediated cancer; ovarian cancer;
KW   ras-mediated cancer; HER/neu promoter; anti-transformation therapy;
KW   anti-cancer therapy; cytostatic; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FI   CDS             1..1389
FI                   /*tag   a
FT                   /product  "Human polyomavirus enhancer activator 3"
XX
PN   WO200004153-A2.
XX
PD   27-JAN-2000.
XX
PF   15-JUL-1999;   99WO-US16142.
XX
PR   15-JUL-1998;   98US-0116049.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Hung M;
XX
DR   WPI; 2000-171269/15.
DR   P-PSDB; AAY44725.
XX
PT   Repression of cell transformation used to suppress tumor growth,
PT   comprises contacting the cell with human polyomavirus enhancer
PT   activator 3 -
XX
PS   Disclosure; Page 89; 92pp; English.
XX
CC   The patent discloses methods for repressing transformation in a cell by
CC   contacting it with polyomavirus enhancer activator 3 (PEA3) to inhibit a
CC   transformed phenotype resulting in reduced tumorigenic or metastatic
CC   potential of a cell. This is used in the treatment of various forms of
CC   cancer, e.g. neu- or ras- mediated cancers. The nucleic acid is
CC   introduced into the mammal through a vector or liposomal complex having
CC   DOTMA, DOPE or DC-Chol. The present cDNA sequence encodes human PEA3.
CC   Human PEA3 binds to a DNA element in the HER/neu promoter and represses
CC   transcription in HER2/neu overexpressing ovarian cancer cells. This can
CC   be used in combination with anti-transformation/anti-cancer therapies or
CC   chemotherapeutic agents.
XX
SQ   Sequence 2064 BP; 427 A; 648 C; 569 G; 420 T; 0 other;


alignment_scores:
            Quality:  230.50              Length:     357
              Ratio:  1.432                 Gaps:      21
 Percent Similarity:  48.459   Percent Identity:  28.571

alignment_block:
 US-09-126-945B-2 x AAZ50016  ..

 Align seq 1/1  to: AAZ50016  from: 1  to: 2064

      44 SerProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLe  60
         |||   :::|||||          :::::|||   ::::::::|||
     265 AGCAGGAAGCCGCCACTCCCCTACCACCATGGCGAGCAGTGCCTTTACTC  314

      60 nSerTyrProAspMetGlyTyrProGluAspSerSerTrpAlaAlaGlyA  77
         |||   :::|||          |||::::::   :::      :::   |
     315 CAGTGCCTATGACCCC......CCCAGACAAATCGCCATCAAGTCCCCTG  358

      77 laProGlyAla......SerSerArgGluGluProProGluGluProGlu  91
         ||||||||||||      :::|||      :::|||      :::   |||
     359 CCCCTGGTGCCCTTGGACAGTCGCCCCTACAGCCCTTTCCCCGGGCAGAG  408
```

```
 92 GlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuVa 108
    |||      :::   |||          ::||||::   :::
409 CAACGGAATTTCCTGAGATCC.....TCTGGCACCTCCCAGCCCCA 449

108 lProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerM 125
     ||||||         |||   ||||||||||
450 CCCTGGCCATGGGTACCTCGGGGAACATAGC..................... 480

125 etValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeu 141

480 ..................................................... 480

142 AsnIleThrAlaAspProMetAspTrpSerProSerAsnValGluLysTr 158
    ::::::      :::|||:::|||         |||
481 TCCGTCTTCCAGCAGCCCCTGGACATTGCCACTCC.............. 516

158 pLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysALaP 175
          :::|||:::::::                |||:::   :
517 .......TTCACATCTCAGGGAGGGGCCGGGAACCCCTCCCAGCCCCCT 559

175 hoGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPhe 191
    ::|||:::      ::::::                ::::::::|||
560 ACCAACACCAGCTGTCCGCAGCCCTGCCCACCCTATCCCCAGCAGAGCTTT 609

192 ArgGln......ArgSerProLeu..................... 197
    :::|||      ::::::|||||
610 AAGCAAGAATACCATGATCCCCTGTATGAACAGGCGGGCCAGCCAGCCGT 659

198 .......GlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerA 212
          ||||||      ::::::::::|||        :::   :::|
660 GGACCAGGGTGGG......GTCAATGGGCACAGG.....TACCCAGGGG 697

212 la.AlaTrpMetLysGluArgThr..................... 219
    || :::|||   ::::::|||:::
698 CGGGGGTGGTGATCAAACAGGAACAGACGGACTTCGCCTACGACTCAGAT 747

220 .SerProGlyAlaIleHisTyrCysAlaSerThrSerGluGlnSer.Trp 235
     ||||||||||||   |||   |||:::||||||:::::   ||| |||
748 GTCACCGGGTGCG...CATCAATGTACCTCCACACAGAGGGCTTCTCTGG 794

236 ThrAspSerGlu..................................... 239
    :::   |||:::
795 GCCCTCTCCAGGTGACGGGGCCATGGGCTATGGCTATGAGAAACCTCTGC 844

240 .......Val.AspSerSerCys...................... 244
              ||:::   |||
845 GACCATTCCCAGATGATGTCTGCGTTGTCCCTGAGAAATTTGAAGGAGAC 894

245 ..........................SerGlyGlnPro..... 248
                            ::|||   |||
895 ATCAAGCAGGAAGGGGTCGGTGCATTCGAGAGGGCCGCCCTACCAGCG 944

249 ..........IleHisLeuTrpGlnPheLeuLysGluLeuLeuLysP 262
            :::::|||||||||||||||      ||||||      |
945 CGGGGTGCCCTGCAGCTGTGGCAATTTCTGGTGGCCTTGCTGGATGACC 994

262 roHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
    ||   :::   ::::::|||||   |||   :::       :::|||:::
995 CA...ACAAATGCCCATTTCATTGCCTGGACGGGC......CGGGGAATG 1045

279 ...PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleAr 294
       |||||:::   :::       :::|||||||||||||||||||||::
1046 GAGTTCAAGCTCATTGAGCCTGAGGAGGTCGCCCAGGCTCTGGGGCATCCA 1095

294 qLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgC 311
    :|||||||||||||||||||||||||||||||||||||||||||:::|||
1096 GAAGAACCGGCCAGCCATGAATTACGACAAGCTGAGCCGCTCGCTCCGAT 1145

311 lnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeu 327
    ||||||:::|||||||||:::::::|||        :::::|||
1146 ACTATTATGACAAAGGCATCATCCAGAAGGTG...CCTGGTGAGCGTTAC 1192

      328 ValTyrGluPheVal 332
          |||||||:::|||||
     1193 GTGTACAAGTTTGTG 1197

Seq_name: [illegible]

seq_documentation_block:
ID   AAQ55149 standard; cDNA to mRNA; 2073 BP.
XX
AC   AAQ55149;
XX
DT   11-JUL-1994  (first entry)
XX
DE   Adenovirus E1A-F gene.
XX
KW   Adenovirus; cancer; ets oncogene; HeLa cell; enhancer core sequence;
KW   methylation; ds.
XX
OS   Human adenovirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             844..1311
FT                   /*tag   a
FT                   /note= "Claimed sequence"
XX
PN   JP05328975-A.
XX
PD   14-DEC-1993.
XX
PF   02-JUN-1992;   92JP-0165453.
XX
PR   02-JUN-1992;   92JP-0165453.
XX
PA   (TAKT ) TAKARA SHUZO CO LTD.
XX
DR   WPI; 1994-021923/03.
DR   P-PSDB; AAK45451.
XX
PT   Novel E1A-F gene  for production of adenovirus E1A-F and cancer
PT   research
XX
PS   Claim 1; Page 6; 7pp; Japanese.
XX
CC   The adenovirus E1A-F gene contains a 474bp open reading frame. The
CC   clone comprising the coding sequence was isolated by screening
CC   a HeLa cell cDNA library.
XX
SQ   Sequence 2073 BP; 458 A; 635 C; 561 G; 418 T; 1 other;


alignment_scores:
            Quality:  230.50              Length:     357
              Ratio:  1.332                 Gaps:      21
 Percent Similarity:  48.459   Percent Identity: 28.571

alignment block:
 US-09-126-945B-2 x AAQ55149  ..

 Align_seq 1/1  to: AAQ55149  from: 1  to: 2073

      44 SerProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLe 60
         |||   ::::|||||          :::::::|||   ::::::::::|||
     187 AGCAGGAAGCCGCCACTCCCCTACCACCATGCCGAGCAGTGCCTTTACTC 236

      60 nSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysA 77
          |||   ::::||          |||:::::::   :::       :::   |
     237 CAGTGCCTATGACCC.....CCCAGACAAATCGCCATCAAGTCCCTG 280

      77 laProGlyAla......SerSerArgGluGluProProGluGluProGlu 91
         ||||||||||||      ::::||      ::::||      :::   |||
     281 CCCCTGGTGCCCCTGGACAGTCGCCCCTACAGCCCTTTCCCCGGGCAGAG 330

      92 GlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuVa 108
```

```
331 CAACGGAATTCCTGAGATCC.....TCTGGCACCTCCCAGCCCCA 371

108 IProGlyGlyLeuThrLeuGluGluHisSerLeuGlnValGlnSerM 125

372 CCCTGGCCATGGGTACCTCGGGGAACATAGC..................... 402

125 etValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeu 141

402 ..................................................... 402

142 AsnIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysIr 158

403 ICCGICTTCCAGCAGCCCCTGGACATTTGCCACTCC.............. 438

158 pLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaP 175

439 .......TTCACATCTCAGGGAGGGGCCGGGAACCCCTCCCAGCCCCCT 481

175 heGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPhe 191

482 ACCAACACCAGCTCTCCGAGCCCTGCCCACCCIATCCCCAGCAGAGCTTT 531

192 ArqGln......ArqSerProLeu..................... 197

532 AAGCAAGAATACCATGATCCCCTGTATGAACAGGCGGGCCAGCCAGCCGI 581

198 .......GlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerA 217

582 GGACCAGGGTGGG.....GTCAATGGGCACAGG.....TACCCAGGGG 619

212 la.AlaTrpMetLysGluArqThr..................... 219

620 CGGGGGTGGTGATCAAACAGGAACAGACGGACTTCGCCTACGACTCAGAT 669

220 .SerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSer.Trp 235

670 GTCACCGGGTGCG...CATCAATGTACCTCCACACAGAGGGCTTCTCTGG 716

236 ThrAspSerGlu.................................. 239

717 GCCCTCTCCAGGTGACGGGGCCATGGGCTATGGCTATGAGAAACCTCTGC 766

240 .......Val.AspSerSerCys..................... 244

767 GACCATTCCCAGATGATGTCTGCGTTGTCCCTGACAAATTTCAAGGAGAC 816

245 .........................SerGlyGlnPro..... 248

817 ATCAAGCAGGAAGGGGTCGGTGCATTTCGAGAGGGGCCGCCCTACCAGCC 866

249 ..........IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysP 262

867 CCGGGGTGCCCTGCAGCTGTGGCAATTTCTGGTGGCCTTGCTGGATGACC 916

263 rsHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGlyLysGlyIle 279

917 CA...ACAAATGCCCATTTCATTGCCTGGACGGCC.....CGGGGAATG 957

279 ...PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleAr 294

958 GAGTTCAAGCTCATTGAGCCTGAGGAGGTCGCCAGGCTCTGGGGCATCCA 1007

294 qLysAsnArqProAlaMetAsnTyrAspLysLeuSerArqSerIleArqS 311

1008 GAAGAACCGGCCAGCCATGAATTACGACAAGCTGAGCCGCTCGCTCCGAT 1057

311 InTyrTyrLysLysGlyIleIleArqLysProAspIleSerGlnArqLeu 327

1058 ACTATTATGAGAAAGGCATCATGCAGAAGGTG...GCTGCTGAGCGTTAC 1104

328 ValTyrGlnPheVal 332

    1105 GTGTACAAGTTTGTG 1119
```


```
seq_documentation_block:
ID   AAV32688 standard; cDNA; 2410 BP.
XX
AC   AAV32688;
XX
DT   20-OCT-1998  (first entry)
XX
DE   Polyomavirus PEA3 cDNA.
XX
KW   Polyomavirus enhancer activator; PEA3; tumour; suppressor; inhibitor;
KW   transformation; HER-2; neu promoter; metastasis; cancer; ss.
XX
OS   Polyomavirus.
XX
FH   Key             Location/Qualifiers
FT   CDS             126..1793
FT                   /*tag   a
FT                   /product   PEA3
FT                   /note   "polyomavirus enhancer activator"
XX
PN   WO9840585-A2.
XX
PD   16-JUL-1998.
XX
PF   12-JAN-1998;   98WO-US00880.
XX
PR   10-JAN-1997;   97US-0780835.
XX
PA   (IEXA ) UNIV IEXAS SYSTEM.
XX
PI   Hung M,  Xing X;
XX
DR   WPI; 1998-399062/34.
DR   P-PSDB; AAW49010.
XX
PT   Method for repressing transformation of cells  by contacting cell
PT   with polyoma virus enhancer activator 3, useful for, e.g. treating
PT   or preventing cancer, tumourigenesis and metastasis
XX
PS   Disclosure; Page 69-70; 83pp; English.
XX
CC   This sequence encodes a polyomavirus enhancer activator, PEA3. This
CC   sequence is used in a method for repressing transformation in a cell
CC   which involves contacting the cell with PEA3 to inhibit a transformed
CC   phenotype. This sequence can also be used in a method to suppress the
CC   growth of a tumour in a mammal comprising introducing to the mammal a
CC   PEA3-encoding nucleic acid where the expression of PEA3 in the mammal
CC   results in a decrease in the growth rate of the tumour. PEA3 can regulate
CC   the HER-2/neu promoter by transcriptional repression and acts as a tumour
CC   repressor. The PEA3 can be used for reducing a transforming,
CC   tumourogenic or metastatic potential of a cell. It can be used for the
CC   prevention and treatment of such transformation driven events as cancer,
CC   tumourigenesis and metastasis.
XX
SQ   Sequence 2410 BP; 521 A; 744 C; 645 G; 500 T; 0 other;
```

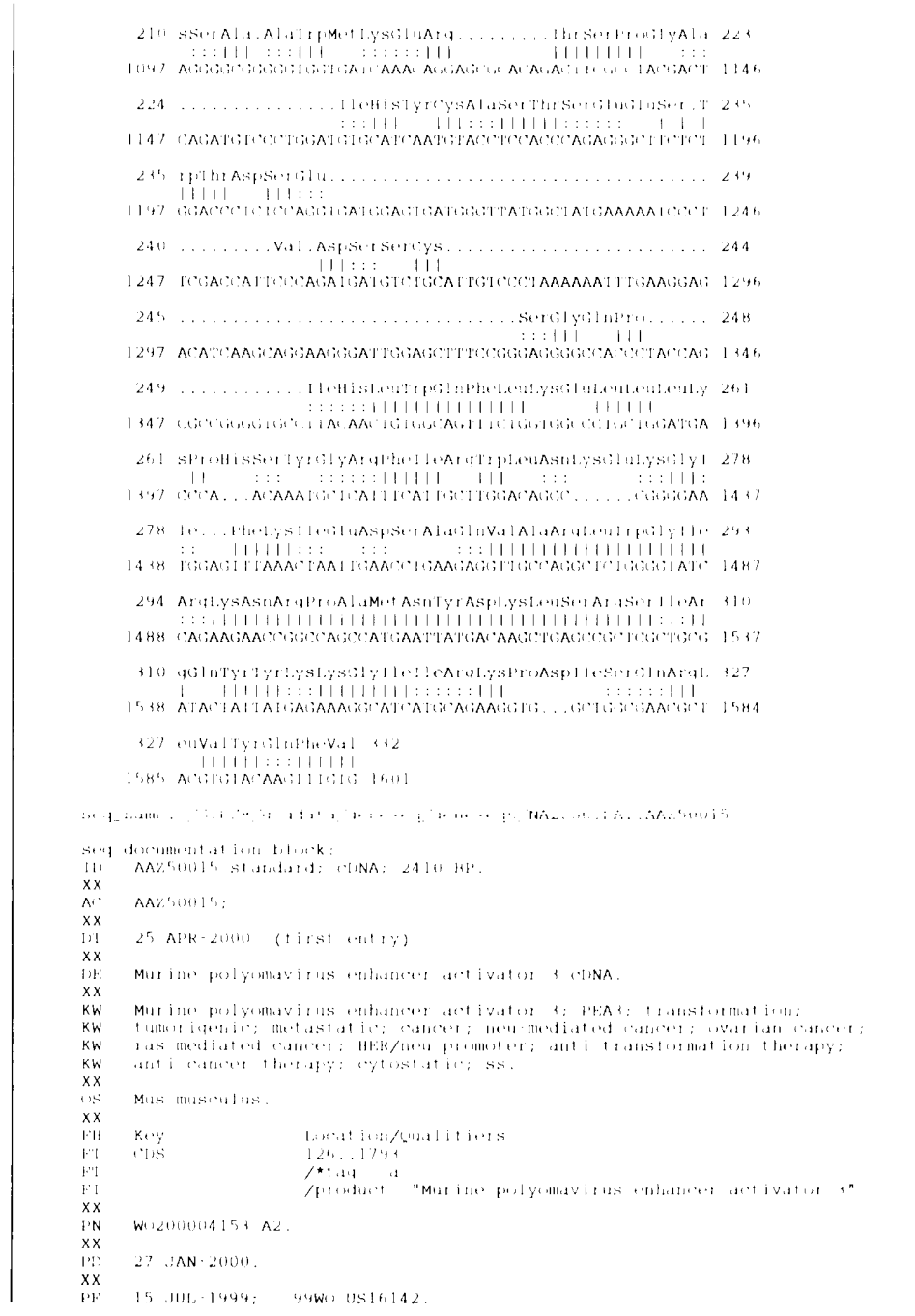

```
alignment_scores:
            Quality:  229.50              Length:    491
              Ratio:  1.153               Gaps:       26
 Percent Similarity:  40.530   Percent Identity: 24.440

alignment_block:
 US-09-126-945B-2 x AAV32688   ..

 Align seq 1/1  to: AAV32688  from: 1  to: 2410

      12 SerProSerHisLeuLeuProProAspThrValSerArqThr..... 26
```





```
XX
PR   15-JUL-1998;   98US-0116049.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Hung M;
XX
DR   WPI; 2000-171269/15.
DR   P-PSDB; AAY44724.
XX
PT   Repression of cell transformation used to suppress tumor growth,
PT   comprises contacting the cell with human polyomavirus enhancer
PT   activator 3 .
XX
PS   Disclosure; Page 86-87; 92pp; English.
XX
CC   The patent discloses methods for repressing transformation in a cell by
CC   contacting it with polyomavirus enhancer activator 3 (PEA3) to inhibit a
CC   transformed phenotype resulting in reduced tumorigenic or metastatic
CC   potential of a cell. This is used in the treatment of various forms of
CC   cancer, e.g. neu- or ras- mediated cancers. The nucleic acid is
CC   introduced into the mammal through a vector or liposomal complex having
CC   DOTMA, DOPE or DC-Chol. The present cDNA sequence encodes murine PEA3.
CC   Murine PEA3 binds directly to HER/neu promoter and represses
CC   transcription in HER2/neu overexpressing ovarian cancer cells. This can
CC   be used in combination with anti-transformation/anti-cancer therapies or
CC   chemotherapeutic agents.
XX
SQ   Sequence 2410 BP; 521 A; 744 C; 645 G; 500 T; 0 other;


alignment_scores:
            Quality:  229.50              Length:    491
              Ratio:  1.153                 Gaps:     26
 Percent Similarity:  40.530   Percent Identity: 24.440

alignment_block:
 US-09-126-945B-2 x AAZ50015   ..

 Align_seq 1/1  to: AAZ50015  from: 1  to: 2410

      12 SerProSerHisLeuLeuLeuProProAspThrValSerArgThr..... 26
         ||||||:::|||          ||||||   :::   :::::|||
     243 TCCCCCGGGCATGCTGAGCATCCCCCCGCGGCTCCGGCACAGACGCCCGG 292

      27 .......................GlyLeuGluLysAlaAlaAlaGlyA 45
                                 |||          :::|||:::::
     293 ACCTCAGGTCTCTGCCTCCGCGCGGGGGCCCGGCCCTGTGGCCGGAGGGA 342

      45 laValGlyLeuGluArgArg............................. 41
         ::       :::|||||||||
     343 GCGGCCGGATGGAGCGGAGGATGAAAGGCGGATACTTGGACCAGCGAGTG 392

      41 ..................................................  41

     393 CCCTACACTTTCTGCAGCAAATCTCCCGGAAATGGGAGCTTGGGCGAAGC 442

      42 ............................AspTrpSerProSerProProA 49
                                     |||   :::     ||||||:
     443 GCTGATGGTCCCGCAGGGAAAGCTCATGGACCCGGGCTCCCTGCCGCCTT 492

      49 laThrProGlu.........GlnGlyLeuSerAlaPheTyrLeuSerTyr 62
         ::      |||        |||   ||||||   |||
     493 CCGACTCAGAAGATCTCTTCCAGGACCTCAGTCACTTC............ 530

      63 PheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla..... 77
                             ::::::::::||   |||::::|||
     531 .....................CAAGAGACGTGGCTCGCAGAAGCTCAGGT 559

      77 ..................................................  77

     560 ACCGGACAGTGATGACAGTTTGTTCCTGATTTCCATTCAGAAAACTTAG 609

      77 ..................................................  77

     610 CTTTCCATAGCCCCACCACAGGATCAAGAAGGAACCCCAGAGTCCCGGC 659

      78 ......ProGlyAlaSerSerArgGluGluProPro.........GluGl 89
               ||||::  |||     :::::||||||          :::::
     660 ACAGACCCCGCCCTGTCCTGCAGCAGGAAGCCACCACTCCCCTACCACCA 709

      89 uProGluGlnCys..............ProValIleAspSerGlnAlaP 101
         :  ||||||||                 :::|||  |||  ||||
     710 TGGAGAGCAGTCCCTTACTCCAGACAAATCGCCATCAAGTCCCCCGCTC 759

     101 roAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHis 117
         |||::                 :::|||||         ||||::
     760 CCGGT..............GCCCCTGGACAGTCGCCCCTGCAGCCCTTT 794

     118 Ser...LeuGluGlnValGlnSerMetValValGlyGluValLeuLysAs 133
         |||      ||||||  |||||:::::  ::::::     :::::
     795 TCCAGGGCAGAACAGCAGCAGAGCCTCCTGAGACGCTCCAGCTCTTCCCA 844

     133 pIleGluThrAlaCysLysLeuLeu.............AsnIleThrAlaA 146
         :  :::  :::          |||             ::::::      :
     845 GTCCCACCCTGGCCACGGGTACCTTGGTGAGCACAGCTCCGTCTTCCAGC 894

     146 spProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThr 162
         ::|||:::|||
     895 AGCCCGTGGACATGTCC................................. 911

     163 GluHisGluLysArgLeuPro.................ProMetGlyLy 178
         |||::::::      |||                   ||||:
     912 ...CACTCCTTCACATCTCCTCAGGGAGGGGGCCGGGAACCTCTCCCAGC 958

     179 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluG 190
         :::|||:::      ::::::                 :::::::
     959 CCCCTATCAACACCAACTGTCCGACCCCTGCCCACCCTACCCCCAGCAGA 1008

     190 lnPheArgGlnArg..................................... 194
         ::|||:::|||:::
    1009 ACTTCAAGCAGGAGTACCATGACCCCCTGTACGAACAGGCTGGCCAGCCC 1058

     195 ...SerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleIrpLy 210
            |||      ||||||         |||               :::
    1059 GCTTCAAGCCAGGGTGGGGTCAGTGGGCACAGG.............TACCC 1096

     210 sSerAla.AlaIrpMetLysGluArg.........ThrSerProGlyAla 223
         :::||| :::|||   :::::::||         ||||||||   :::
    1097 AGGGGCCGGGGTGGTGATCAAACAGCAGCGCACAGACTTCGCCTACGACT 1146

     224 ..............IleHisTyrCysAlaSerThrSerGluGlnSer.T 235
                       :::|||   |||:::|||||:::::   ||| |
    1147 CAGATGTCCCTGGATGTGCATCAATGTACCTCCACCCAGAGGGCTTCTCT 1196

     235 rpThrAspSerGlu..................................... 239
         |||||   |||::
    1197 GGACCCTCTCCAGCTGATGGAGTGATGGTTATGGCTATGAAAAATGCCT 1246

     240 .........Val.AspSerSerCys........................ 244
                  ||:::   |||
    1247 TCGACCATTCCCAGATGATGTCTGCATTGTCCCTAAAAAATTTGAAGGAG 1296

     245 ..............................SerGlyGlnPro...... 248
                                      :::|||   |||
    1297 ACATCAAGCAGGAAGGGATTGGAGCTTTCCGGGAGGGGCCACCCTACCAG 1346

     249 ............HisHisLeuIrpGlnPheLeuLysGluLeuLeuLeuly 261
                    ::::::||||||||||||||||      ||||||
    1347 CGCCGGGGTGCCTTACAACTGTGGCAGTTTCTGGTGGCCCTGCTGGATGA 1396

     261 sProHisSerTyrGlyArgPheIleArgIrpLeuAsnLysGluLysGlyI 278
         |||   :::   ::::::||||||'   |||   :::      :::|||:
    1397 CCCA...ACAAATGCTCATTTCATTGCTTGGACAGGC......CGGGGAA 1437

     278 le...PheLysIleGluAspSerAlaGlnValAlaArgLeuIrpGlyIle 293
```

```
          ::   |||||||:::   :::      :::|||||||||||||||||||||
     1438 TGGAGTTTAAACTAATTGAACCTGAAGAGGTTGCCAGGCTCTGGGGTATC 1487

      294 ArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleAr 343
          :::||||||||||||||||||||||||||||||||||||||||||||||||
     1488 CAGAAGAACCGGCCAGCCATGAATTATGACAAGCTGAGCCGCTCGCTGCC 1537

      310 qGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgL 327
          :   ||||||:::|||||||||||:::::|||      :::::|||
     1538 ATACTATTATGAGAAAGGCATCATGCAGAAGGTG...GCTGGCGAACGCT 1584

      327 qaValTyrGlnPheVal 332
            ||||||:::||||||
     1585 ACGTGTACAAGTTTGTG 1601

seq_name: /S1s/s2sqdata/geneseq-geneseqpzNAZ/pzNA_AAF54019

seq documentation block:
ID   AAF54019 standard; cDNA; 2410 BP.
XX
AC   AAF54019;
XX
DT   30-MAR-2001  (first entry)
XX
DE   Mouse PEA 3 protein encoding cDNA, SEQ ID NO:13.
XX
KW   Age related gene regulation; gene expression; human protein C; hPC;
KW   5' UTR; 5' untranslated region; age regulatable expression construct;
KW   PEA 3 element; polyoma virus activator 3; antisense therapy;
KW   gene therapy; thrombosis; cardiovascular disease; diabetes;
KW   Alzheimer's disease; Parkinson's disease; cancer; osteoporosis;
KW   osteoarthritis; dementia; ss.
XX
OS   Mus musculus.
XX
PN   WO200075279-A2.
XX
PD   14-DEC-2000.
XX
PF   06-JUN-2000; 2000WO-US15728.
XX
PR   09-JUN-1999;  1999US-0328925.
XX
PA   (UNMI-) UNIV MICHIGAN.
XX
PI   Kurachi K, Kurachi S;
XX
DR   WPI; 2001-061708/07.
DR   P-PSDB; AAB60290.
XX
PT   New regulatory elements that control age related gene expression,
PT   useful in gene therapy and for reducing Factor IX expression
XX
PS   Disclosure; Fig 9A; 225pp; English.
XX
CC   The invention relates to nucleic acid sequences which regulate gene
CC   expression in an age related manner and/or in a liver specific manner.
CC   The invention identifies regions of the human factor IX (hFIX) gene, and
CC   a region of the human protein C (hPC) gene, which are age-related
CC   regulatory sequences. The hFIX age related regulatory sequences are
CC   designated AE5' (AAF54016) and AE3' (AAF54017) and are found in the 5'
CC   UTR (at position 2164-2165 of AAF54018) and 3' UTR (at position
CC   34383-45655 of AAF54018) respectively. These elements act synergistically
CC   to increase hFIX levels over the lifespan of an individual; however, they
CC   can independently exert effects on hFIX mRNA in an age related manner,
CC   with AE5' acting to stabilise hFIX mRNA, and AE3' acting to increase hFIX
CC   mRNA levels, over time. AE5' also directs liver specific expression. The
CC   hPC gene age-related regulatory sequence is found in the 5' UTR
CC   (AAF54081), and contains two PEA 3 (polyoma virus activator 3) elements
CC   5' GAGGAAA 3' and 5' CAGGAAG 3'. The age-related regulatory sequences of
CC   the invention, along with their homologues, variants and fragments, may
CC   be used in the construction of recombinant expression vectors for the
CC   expression of a desired sequence in an age related fashion in a host
CC   cell. Preferred target genes for expression in such age regulatable
CC   expression vectors include blood-borne therapeutic proteins and proteins
CC   involved in blood coagulation (e.g., the pro-coagulants factor IX and factor VIII, and the
CC   anti-coagulants protein C and antithrombin III), human
CC   alpha 1-antitrypsin, PEA-3 protein and reporter proteins such as
CC   luciferase. Preferred promoters for use in such age regulatable
CC   expression vectors include the human factor IX promoter, the T7 promoter,
CC   the T3 promoter and the SP6 promoter. The expression vectors of the
CC   invention may be used in gene therapy to provide age- related and/or
CC   liver specific expression of target genes. Age-regulatable constructs may
CC   be used in the treatment of such age-related conditions such as
CC   thrombosis, cardiovascular disease, diabetes, Alzheimer's disease,
CC   Parkinson's disease, cancer, osteoporosis, osteoarthritis and dementia.
CC   Specifically, they may be used to express factor IX antisense mRNA in the
CC   treatment of thrombotic conditions associated with the natural
CC   age related rise in factor IX expression. Transgenic cells or animals
CC   that contain vectors of the invention are useful as models of these
CC   diseases, in screening for potential therapeutic agents and for studying
CC   normal processes such as ageing and gene expression. Fragments and
CC   homologues of age related regulatory sequences, are useful as probes to
CC   detect, isolate or identify other such sequences in samples. The present
CC   sequence represents a nucleic acid sequence which may be incorporated
CC   into a vector of the invention.
XX
SQ   Sequence 2410 BP; 521 A; 744 C; 645 G; 500 T; 0 other;


alignment scores:
            Quality:  229.50              Length:    491
              Ratio:  1.153                 Gaps:     26
  Percent Similarity:  40.530   Percent Identity: 24.440

alignment block:
 US-09-126-945B-2 x AAF54019   ..

 Align seq 1/1  to: AAF54019  from: 1  to: 2410

      12 SerProSerHisLeuLeuLeuProProAspThrValSerArgThr..... 26
         ||||||:::|||          ||||||   :::   :::::|||
     243 PCCCCGGGCATGCTGAGCATCCCCCGGCGGCTCCGGCACAGACGCCCGG 292

      27 .........................GlyLeuGluLysAlaAlaAlaGlyA 45
                                 |||         :::|||:::|||
     293 ACCTCAGGTCTCTGCCTCCGCGCGGGGGCCCGGCCCTGTGGCCGGAGGGA 342

      46 IaValGlyLeuGluArgArg..................................... 41
         ::    :::||||||||||
     343 GCGGCCGGATGGAGCGGAGGATGAAAGGCGGATACTGGACCAGCGAGTG 392

      41 ......................................................... 41

     393 CCCTACACCTTCTGCAGCAAATCTCCCGGAAATGGGAGCTTGGGCGAAGC 442

      42 ..........................AspTrpSerProSerProProA 49
                                  |||   :::       ||||||
     443 GCTGATGGTCCCGCAGGGAAAGCTCATGGACCCGGGTCCCTGCCGCCTT 492

      49 IaIleProGln.........GlnGlyLeuSerAlaIleTyrLeuSerTyr 82
         ::      |||         |||   ||||||   |||
     493 CCGACTCAGAAGAICTCTTCCAGGACCTCAGTCACTTC............ 530

      83 PheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla..... 77
                             ::::::::::|||   |||:::|||
     531 ..................CAAGAGCGTGGCTCGCAGAAGCTCAGGT 559

      77 ......................................................... 77

     560 ACCGGACAGTGATGAGCAGTTTGTTCCTGATTCCATTCAGAAAACTTAG 609

      77 ......................................................... 77

     610 CTTTCCATAGCCCCACCACCAGCATCAAGAAGGAACCCCAGAGTCCCCGC 659
```

```
 78 ......ProGlyAlaSerSerArgGluGluProPro.........GluGl 89
          ||||||    |||      ::::::||||||          :::::
660 ACAGACCCTGCCCTGTCCTGCAGCAGGAAGCCACCACTCCCCTACCACCA 709

 89 qProGluGlnCys..............ProValIleAspSerGluAlaP 101
     :   ||||||||               ::|||   |||   ||||
710 TGGAGAGCAGTGCCTTTACTCCAGACAAATCGCCATCAAGTCCCCCCCTC 759

101 roAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHis 117
    |||::                 :::||||||         ||||::
760 CCGGT..............GCCCCTGGACAGTCGCCCCTGCAGCCCTTT 794

118 Ser...LeuGluGlnValGlnSerMetValValGlyGluValLeuLysAs 133
    |||      ||||||   ||||||:::::   ::::::     :::::
795 TCCAGGGCAGAACAGCAGCAGAGCCTCCTGAGAGCCTCCAGCTCTTCCCA 844

134 pIleGluThrAlaCysLysLeuLeu.............AsnIleThrAlaA 146
    :  :::  :::        |||              ::::::        :
845 GTCCCACCCTGGCCACGGGTACCTTGGTGAGCACAGCTCCGTCTTCCAGC 894

146 spProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThr 162
    ::|||:::|||
895 AGCCCGTGGACATGTGC............................... 911

163 GluHisGlnTyrArgLeuPro.................ProMetGlyLy 178
    ||||:::::      |||                  ||||::
912 ...CACTCCTTCACATCTCCTCAGGGAGGGGCCGGGAACCTCTCCCAGC 958

178 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGlnG 190
      :::|||:::      ::::::                 :::::::
959 CCCCTATCAACACCAACTGTCGGAGCCCTGCCCACCCTACCCCCAGCAGA 1008

190 InPheArgGlnArg.................................. 194
    ::|||::::|||:::
1009 ACTTCAAGCAGGAGTACCATGACCCCCTGTACGAACAGCCTGGCCAGCCC 1058

195 ...SerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLy 211
       |||      ||||||          |||              :::
1059 GCTTCAAGCCAGGGTGGGGTCAGTGGGCACAGG.............TAGC 1096

210 sSerAla.AlaTrpMetLysGluArg.........ThrSerProGlyAla 223
     :::||| :::|||   :::::::|||         |||||||||   :::
1097 AGGGGCGGGGTGGTGATCAAACAGGAGGCGCACAGACTTCGGCTACGACT 1146

224 ..............IleHisTyrCysAlaSerThrSerGluGluSer.I 235
                 :::|||   |||::::|||||||||::::   |||  |
1147 CAGATGTCCCTGGATGTGCATCAATGTAGCCTCCACCCAGAGGGCTTCTCT 1196

235 rpThrAspSerGlu                              ...  .. 239
    ||||||   |||::                                
1197 GGACCCTCTCCAGGTGATGGAGTGATGGGTTATGGCTATCAAAAATCCCT 1246

240 .........Val.AspSerSerCys..................... 244
                |||::   |||
1247 TCGACCATTCCCAGATGATGTCTGCATTGTCCCTAAAAAATTTGAAGGAG 1296

245 ..............................SerGlyGluPro..... 248
                                  ::|||   |||
1297 ACATCAAGCAGGAAGGGATTGGAGCTTTCCGGGAGGGGCCAGCCTATCCAG 1346

249 ............IleHisLeuTrpGlnPheLeuLysGluGlnSerGly 261
              ::::::||||||||||||||||      ||||||
1347 CGCCGGGGTGCCTTACAACTGTGGCAGTTTCTGGTGGCCCTGCTGGATGA 1396

261 sProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluGlyGly 278
    |||   :::   ::::::||||||   |||   :::      ::::|||
1397 CCCA...ACAAATGCTCATTCATTGCTTGGACAGGC.....CGGGGAA 1437

278 Ie...PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIle 293
    ::   ||||||::   :::      ::|||||||||||||||||||||||
1438 TGGAGTTTAAACTAATTGAACCTGAAGAGGTTGCCAGGCTCTGGGGTATC 1487

294 ArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleAr 416
```

```
      :::|||||||||||||||||||||||||||||||||||||||||:::||
1488 CAGAAGAACCGGCCAGCCATGAATTATGACAAGCTGAGCCGCTCGCTGCG 1537

 310 qGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgL 327
     |   ||||||:::|||||||||:::::|||         :::::|||
1538 ATACTATTATGAGAAAGGCATCATGCAGAAGGTG...GCTGGCGAACGCT 1584

 327 euValTyrGlnPheVal 332
       ||||||:::|||||
1585 ACGTGTACAAGTTTGTG 1601

seq_name: /SIOSE/[illegible]/NAZ0001.PAT:AAZ94395

seq_documentation_block:
ID   AAZ94395 standard; cDNA; 637 BP.
XX
AC   AAZ94395;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Human ELF5 short transcript cDNA.
XX
KW   ELF5 short transcript; human; ETS protein; transcription factor;
KW   cell proliferation; cell differentiation; therapy; diagnosis;
KW   cancer; tumour; autoimmune disease; chromosome 11p13 15; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             134..616
FT                   /*tag= a
XX
PN   WO200012695-A1.
XX
PD   09-MAR-2000.
XX
PF   26-AUG-1999;    99WO-AU00691.
XX
PR   27-AUG-1998;    98AU-0005512.
PR   30-SEP-1998;    98AU-0006252.
XX
PA   (MONU ) UNIV MONASH.
XX
PI   Kola I,  Zhou J;
XX
DR   WPI; 2000-270820/23.
DR   P-PSDB; AAY79289.
XX
PT   ELF5 protein, useful for diagnosis, antibody generation and as
PT   screening tool for agents capable of modulating transcriptional events
PT   during cellular functioning such as in tumorigenesis, comprises an ELS
PT   domain -
XX
PS   Claim 7; Page 89-90; 114pp; English.
XX
CC   The present sequence of that of cDNA encoding human ELF5 short
CC   transcript (see AAY79288), an isoform of the newly identified ELF5
CC   (see AAY79288).  Human ELF5 and ELF5 short transcript differ in size
CC   due to splicing out of exon regions from the ELF5 mRNA molecule
CC   encoding ELF5 short transcript.  ELF5 is defined as belonging to the
CC   ETS family of transcription factors due to its expression of an
CC   ETS domain which recognises and binds purine-rich GGA(A/T) core
CC   motifs.  The ELF5 gene was mapped to human chromosome 11p13 15.
CC   ELF5 expression was not detectable in 20/20 human breast carcinomas
CC   examined.  The identification of novel human ELF5 permits the
CC   identification and rational design of a range of products for use
CC   in therapy, diagnosis and antibody generation involving, e.g.,
CC   regulation of cellular functional activity such as cellular
CC   differentiation or cellular proliferation.  These therapeutic
CC   molecules may also act as either agonists or antagonists of ELF5
CC   function and will be useful in cancer and autoimmune disease
CC   therapy.
XX
```

```
SQ   Sequence 647 BP; 180 A; 145 C; 157 G; 155 T; 0 other;


alignment scores:
              Quality:  225.50              Length:     129
                Ratio:  2.854                 Gaps:       3
 Percent Similarity:  61.240   Percent Identity:  45.669

alignment block:
 US-09-126-945B-2 x AAZ94495   ..

 Align seq 1/1  to. AAZ94495  from. 1  to. 647

      244 SerTrpThrAsp..................................... 237
          ||||||||||||
      197 TCGTGGACTGATCTGTTCAGCAATGAAGAGTACTACCCTGCCTTTGAGCA 246

      248 ....SerGluValAspSerSerCys........................ 244
              :::::::::|||||||:::|||
      247 TCAGACAGATGCTGATTCCAACTGCTTGAAAACAAGTGGCATCAAAAGCC 296

      245 ........................SerGlyGluProIleHisLeuTrp 262
                                  |||   |||      |||||||||
      297 AAGACTGCCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGC 346

      263 GluPheLeuLysGlnLeuLeuLeuLysProHisSerTyrGlyArgPheTl 269
          :::|||::::::::::|||||||||:::|||:::::          :::::
      347 GAATTTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTT 396

      269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaG 289
          :::|||   ::::::|||:::|||||||||||:::::      |||
      397 GGAATGGGAAGATAGGAACAAGGAATTTTTCGGGTGCTTAAATGGGAAG 446

      286 InValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
          :::|||::::::::|||||   |||||||:::::      |||:::|||
      447 CCCTGGCAAAGATGTGGGGACAAAGGAAGAAAAATCACACAATGACATAT 496

      303 AspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 319
          :::|||||||||||||:::::|||   ||||||||   |||||:::::
      497 GAAAAGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGGAATTTTGCA 546

      319 qLysProAspIleSerGlnArgLeuValTyrGlnPhe 341
          ::::   |||       :::||||||||||||:::|||
      547 GCGGGTTGAC......CGAAGGTTAGTGTACAAATTT 577

seq name: /SIDS8/gcgdata/geneseq/geneseqp/NA1990_DAT:AAQ50662

seq documentation block:
ID   AAQ50662 standard; DNA; 1890 BP.
XX
AC   AAQ50662;
XX
DT   26-MAY-1994  (first entry)
XX
DE   Human Hum Fli-1 gene (genomic DNA).
XX
KW   chromosomal translocation; chimeric; chimaeric; Ewing sarcoma;
KW   Ews gene; malignant melanoma; hum fli-1;
KW   primitive peripheral neuroectodermal tumour; human chromosome 11;
KW   human chromosome 22; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            1..89
FT                   /*tag   a
FT                   /number  1
FT                   /*old_start  22..21
FT   intron          40..99
FT                   /*tag   b
FT                   /number  1
FT   exon            100..311
FT                   /*tag   c
FT                   /number  2
FT   intron          312..371
FT                   /*tag   d
FT                   /number  2
FT   exon            372..526
FT                   /*tag   e
FT                   /number  3
FT   intron          527..586
FT                   /*tag   f
FT                   /number  3
FT   exon            587..790
FT                   /*tag   g
FT                   /number  4
FT   intron          791..850
FT                   /*tag   h
FT                   /number  4
FT   exon            851..916
FT                   /*tag   i
FT                   /number  5
FT   intron          917..976
FT                   /*tag   j
FT                   /number  5
FT   exon            977..1042
FT                   /*tag   k
FT                   /number  6
FT   intron          1043..1102
FT                   /*tag   l
FT                   /number  6
FT   exon            1103..1162
FT                   /*tag   m
FT                   /number  7
FT   intron          1163..1222
FT                   /*tag   n
FT                   /number  7
FT   exon            1223..1270
FT                   /*tag   o
FT                   /number  8
FT   intron          1271..1330
FT                   /*tag   p
FT                   /number  8
FT   exon            1331..1890
FT                   /*tag   q
FT                   /number  9
XX
PN   WO9323549-A.
XX
PD   25-NOV-1993.
XX
PF   19-MAY-1993;   93WO-FR000494.
XX
PR   20-MAY-1992;   92FR-0006123.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Aurias A;  Delattre O;  Desmaze C;  Melot T;  Peter M;
PI   Plouqastel B;  Thomas G;  Zucman J;
XX
DR   WPI; 1993-386580/48.
DR   P-PSDB; AAR44556.
XX
PT   New nucleic acid of EWS gene and its hybrid(s)  - contg. gene
PT   sequence involved in chromosomal trans-location, also derived
PT   mRNA, probes, fusion proteins etc., for diagnosis and treatment
PT   of Ewing sarcoma and melanoma
XX
PS   Disclosure; Fig 7 and Fig 13; 124pp; French.
XX
CC   The intron-exon junctions of the human Ews gene and the Hum Fli-1
CC   gene have been sequenced (see AAQ50646 and AAQ50662, respectively).
CC   The different fusion products which could be formed by fusing
CC   exons from the two genes, as happens after specific chromosomal
CC   translocations, can be predicted (see AAQ50671-Q50678). The sequences
CC   at fusion junctions of other observed translocations are given in
```

CC AAg50675-g50684). NOTE: the 1890 residue sequence AAg50682 does not
CC appear in the specification; it is a combination of the Hum Fli-1
CC cDNA sequence (Fig 7, AAg50644) and the individual intron sequences
CC with their intron-exon junctions (Fig 13).
XX
Sq Sequence 1890 BP; 457 A; 561 C; 458 G; 414 T; 0 other;

Alignment_scores:

Quality:	225.00	Length:	338
Ratio:	1.731	Gaps:	11
Percent Similarity:	48.462	Percent Identity:	23.669

241 spSerSerCys..... 244
:::|||
1280 CCTTGGCCTGAAGCCTTTTTGTTCTCTCCTTCCTCAGCGTGTA 1429

245 ...SerGlyGlnProIleHisLeuTrpGlnPheLeuLysGlnLeuLeuLe 286
|||||||| |||:::|||||||||||||| || ||||
1330 GGAAGGGGCAG...ATCAGCTGTGCAATTCCTCTGGAG...CTGT 1474

286 alysProHisSerTyrGlyArgPheIleArgIleuAsnLysGluLysG 277
|::: ::: ::: ||| || ::: :::|
1374 CTCCGACAGAGCAAGGCAGCTGTATCACCTGG...GAGGGACCAAAG 1420

```
CC   library of virtual compounds in silico according to defined criteria,
CC   and evaluating in silico the binding of the virtual compounds with the
CC   tNA according to defined criteria. Also described are: (1) a method of
CC   defining a set of oligonucleotides (ONs) that modulate the expression of
CC   a tNA sequence via binding of the ONs with the tNA sequence comprising
CC   generating a library of virtual compounds in silico according to defined
CC   criteria, and evaluating in silico the binding of the virtual ONs with
CC   the tNA according to defined criteria; and (2) a method of defining a
CC   set of compounds that modulate the expression of a tNA sequence via
CC   binding of the compounds with the tNA. The methods can be used for the
CC   generation and identification of synthetic compounds having defined
CC   physical, chemical or bioactive properties. Information gathered from
CC   assays of such compounds is used to identify nucleic acid sequences that
CC   are tractable to a variety of nucleotide sequence based technologies,
CC   e.g. antisense drug discovery and target validation. AAZ10852 to
CC   AAZ41220, and AAY52701 to AAY52706, represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 2266 BP; 457 A; 741 C; 605 G; 474 T; 0 other;


alignment scores:
            Quality:  225.00              Length:     168
              Ratio:  2.228                 Gaps:       8
 Percent Similarity:  60.119   Percent Identity:  36.310

alignment block:
 US-09-126-945B-2 x AAZ41049   ..

 Align seq 1/1  to: AAZ41049  from: 1  to: 2266

     169 ProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAl 185
         ||||||   |||   |||       ::::::|||:::
     124 CCGCCCCCCGGAGCTGCC......AACATTGCCAACGCCACGCCCACGCT 167

     185 aMetSerGluGluGlnPheArgGluArgSerProLeuGlyGlyAspValL 202
           :::   :::   |||||:::       |||:::   |||   ::::
     168 ACACACAGCCTCAACTTTCAGGAGA......CCCGTCCGTGGC...CTTA 208

     202 euHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArq 218
         ::::::::   |||   :::   :::|||                 :::
     209 TTTATTCCACCCTTCCTGTACATCGTAGC.................GAA 240

     219 ThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTr 235
         :::     ||||||                         :::     ||
     241 TCAATCCGTGGCGCCGCACTCCTCCGCATCCCTCTTTAACACTACCCCTG 290

     235 p.............ThrAspSerGluValAspSerSerCysSerGlyGluP 248
         |             |||           :::|||   |||
     291 GGATGGCGTGAGCACTCCCCCAGCGATGGACCCATCT.............. 327

     248 roIleHisLeuTrpGlnPheLeuLysGluLeuLeuLysProHisSer 264
           :::   ||||||||||||||||   :::   ||||||:::   ::::::
     328 ..GTGACGCTGTGGCAGTTTCTGCTGCAG...CTGCTGAGAGAGCAAGGC 372

     265 TyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIl 281
         |||::::::|||   |||   :::::::::   |||   ||||||::
     373 AATGGCCACATCATCTCCTGGACTTCACGGATGGTGGTGAATTCAAGCT 422

     281 eGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgP 298
         :   |||:::   :::||||||||||||||||||||:::|||||||:::
     423 GGTGGATGCAGAGGAGGTGGCCCGGCTGTGGGGACTACGCAACAACAAGA 472

     298 roAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLys 314
         |||||||||||||||||||||||||:::::|||   ||||||
     473 CCAACATGAATTACGACAAGCTCAGCCGGGCCTTGCGGTACTACTATGAC 522

     315 LysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPh 331
         |||:::||||||||||   :::   :::|||:::::|||||:::||
     523 AAGAACATCATCCGCAAGGTGAGC...GGCCAGAAGTTCGTCTACAACTT 569

     331 eVal 332
         ||||
```

```
     570 TGTG 573

Seq_Name: /SIDS8-petlab/geneseqp/NA2000.DAT:AAZ06608

seq_documentation_block:
ID   AAZ06608 standard; DNA; 2266 BP.
XX
AC   AAZ06608;
XX
DT   23-NOV-1999  (first entry)
XX
DE   Human ELK-1 nucleotide sequence.
XX
KW   Human ELK 1; p62TCF; Ets domain transcription factor protein; apoptosis;
KW   expression inhibition; infection; inflammation; tumour formation;
KW   diagnosis; phosphorothioate; antisense compound; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             316..1602
FT                   /*tag= a
FT                   /product= "ELK_1
XX
PN   US5948680-A.
XX
PD   07-SEP-1999.
XX
PF   17-DEC-1998;  98US-0213767.
XX
PR   17-DEC-1998;  98US-0213767.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Baker BF, Cowsert LM;
XX
DR   WPI; 1999-517959/43.
DR   P-PSDB; AAY32824.
XX
PT   Antisense compound useful for diagnosis, treatment and prevention of
PT   disease associated with ELK 1 expression
XX
PS   Example 13; Column 41 46; 81pp; English.
XX
CC   This is the human ELK-1 nucleotide sequence. Human ELK 1 also known as
CC   p62TCF is a member of the ternary complex factor subfamily of Ets domain
CC   transcription factor proteins. Antisense polynucleotides targeted to the
CC   ELK 1 nucleic acid molecule AAZ06571 Z06607 inhibit the expression of
CC   ELK-1. Sequences AAZ06571 Z06607 all cause at least 80% inhibition of
CC   ELK-1 expression. The antisense sequences can be used to inhibit the
CC   expression of human ELK 1 in human cells or tissues in vitro. ELK 1 uses
CC   a bipartite recognition mechanism mediated by both protein-DNA and
CC   protein-protein interactions to regulate genes by direct and indirect DNA
CC   binding and has been shown to control various signal transduction
CC   pathways and other cell functions including apoptosis. This means that
CC   antisense compounds inhibiting expression of ELK-1 can be used to treat
CC   diseases associated with its expression in animals, particularly humans
CC   and to prevent or delay infection, inflammation or tumour formation. The
CC   compounds can also be used for diagnosis, as research reagents and in
CC   kits.
XX
SQ   Sequence 2266 BP; 457 A; 741 C; 605 G; 474 T; 0 other;


alignment_scores:
            Quality:  225.00              Length:     168
              Ratio:  2.228                 Gaps:       8
 Percent Similarity:  60.119   Percent Identity:  36.310

alignment_block:
 US-09-126-945B-2 x AAZ06608   ..

 Align seq 1/1  to: AAZ06608  from: 1  to: 2266
```

```
      169 ProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAl 187
          ||||||   |||   |||      :::::::::::
      124 CCGGCCCCGGGAGCTGCC......AACATTGCCAACGCCACCGCCACGCT 167

      188 aMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValL 207
            :::   :::   |||||:::      |||:::   |||   ::::
      168 ACACACAGCCTCAACTTTCAGCAGA.....CCCGTCCGTGGC...CTTA 208

      208 cvsHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGlnAr 227
          ::::::::   |||   :::   :::|||                  :::
      209 TTTATTCCACCCTTCCTGTACATCGTAGC.................GAA 240

      228 ThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTr 247
          :::      ||||||                  :::      ||
      241 TCAATCCGTGGCGCCGCACTCCTCCGCATCCCTCTTTAACAGTACCCCTG 290

      248 p............ThrAspSerGluValAspSerSerCysSerGlyGlnP 268
          |           |||          ::::|||   |||
      291 CGATGGCGTGAGCACTCCCCCAGCGATGGACCCATCT............. 327

      269 roIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSer 288
            :::   ||||||||||||||||   :::   ||||||:::   ::::::
      328 ..GTCAGCTGTGGCAGTTTCTGCTGCAG...CTGCTGAGAGAGCAACGC 372

      289 TyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysTl 308
            :|:::::::::|||   |||   :::::::::   |||   |||||||:
      373 AATGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGAATTCAAGCT 422

      309 eGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgP 328
          :   ||||::   :::||||||||||||||||||||:::|||||||||:::
      423 GGTGGATGCAGAGGACGTGGCCCCGCTCTGCGCACTACGCAAGAACAAGA 472

      329 roAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLys 348
            ||||||||||||||||||||||||:::::::|||   ||||||
      473 CCAACATGAATTACGACAAGCTCAGCCGGGCCTTGCGGTACTACTATGAC 522

      349 LysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPh 368
          |||:::|||||||||||||   :::   ::::|||:::::|||||:::||
      523 AAGAACATCATCCGCAAGGTGAGC...GGCCAGAAGTTCGTCTACAAGTT 569

      369 eVal 372
          ||||
      570 TGTG 573

Seq_name:  ...  ...pdfa  ...  ...seq  ...esp_NA1999-2A:AAZ33671

seq_documentation_block:
ID   AAZ33671 standard; cDNA; 2266 BP.
XX
AC   AAZ33671;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Human breast tumour-associated EST 61.
XX
KW   Expressed sequence tag; EST; human; breast; cancer; gene therapy;
KW   treatment; tumour; cytostatic; medicament; ss.
XX
OS   Homo sapiens.
XX
PN   DE19813839-A1.
XX
PD   23-SEP-1999.
XX
PF   20-MAR-1998;   98DE-1013839.
XX
PR   20-MAR-1998;   98DE-1013839.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E,  Rosenthal A;
XX
DR   WPI; 1999-528981/45.
DR   P-PSDB; AAY48614.
XX
PT   Human nucleic acid sequences and protein products from tumor breast
PT   tissue, useful for breast cancer therapy -
XX
PS   Claim 3; 139; 188pp; German.
XX
CC   This invention decribes novel human nucleic acid sequences from tumor
CC   breast tissue which have cytostatic activity. The nucleic acid sequences
CC   can be used to produce and isolate full-length gene sequences. They can
CC   be used to express proteins, which can be used as tools to find an
CC   activity against breast cancer. The sequences can be used in sense or
CC   antisense form. They are especially useful for medicaments for gene
CC   therapy to treat breast cancer. AAZ33611 Z48617 represents expressed
CC   sequence tags described in the method of the invention.
XX
SQ   Sequence 2266 BP; 526 A; 580 C; 654 G; 506 T; 0 other;


alignment_scores:
            Quality:  209.50              Length:    291
              Ratio:  1.650                 Gaps:     11
 Percent Similarity:  44.643   Percent Identity: 24.055

alignment_block:
 US-09-126-945B-2 x AAZ33671   ..

 Align seq 1/1  to. AAZ33671  from. 1  to. 2266

      148 MetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHi 164
          :::|||||||          :::|||:::|||                  ::
           68 CTGGACTGGATCAGCTACCAAGTGGAGAAG.................AA 99

      164 sGlnTyrArgLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyL 181
          :::|||              :::      :::      ::::::   |||
      100 CAAGTACGACGCAAGGGCCATTGACTTCTCACGATGTGACATGGATGGCG 149

      181 ysGluLeuCysAlaMetSerGlnGlnGlnPheArg...GlnArgSerPro 196
            ||||||      :::   |||::::::|||          :::|||
      150 CCACCCTCTGCAATTGTGCCCTTGAGGAGCTGCGTCTGGTCTTTGGGCCT 199

      197 LeuGlyGlyAspValLeuHisAlaHisLeu..................... 206
          |||   ||||||   ||||||||||:::|||
      200 CTG...GGGGACCAACTCCATGCCCAGCTGCGAGACCTCACTTCCAGCTC 246

      207 .........................AspIle.TrpLysSerAlaAlaTrp 214
                                  ::::::.||:::      ||||||
      247 TTCTGATGAGCTCAGTTGGATCATTGAGCTGCTGGAGAAGGATGGCATGG 296

      215 MetLysGluArg.......................................... 218
          :::::::|||
      297 CCTTCCAGGAGGCCCTAGACCCAGGGCCCTTTGACCAGGGCAGCCCCTTT 346

      219 .....................................ThrSerProG 222
                                              |||:::|||:
      347 GCCCAGGAGCTGCTGGACGACGGTCAGCAAGCCAGCCCCTACCACCCCGG 396

      222 TyAlaIle.............HisTyrCysAlaSerThrSer...Glu 242
          ::|||:::                  ::::||   ||||||
      397 CAGCTGTGGCGCAGGAGCCCCCTCCCCTGGCAGCTCTGACGTCTCCACCG 446

      243 GluSerTrpThrAspSerGluVal.AspSerSerCysSerGly..... 266
          :::::|||      ||||||||:::    ||||||   ||||||
      447 CAGGACTGGTGGTTCTCGGAGCTCCCACTCCTCAGACTCCGGTGGAAGTG 496

      246 ................................................... 246

      497 ACGTGGACCTGGATCCACTGATGGCAAGCTCTTCCGCAGCGATGTTTT 546

      246 ................................................... 246

      547 CGTGACTGCAAGAAGGGGGATCCCAAGCACGGGAAGCGGAAACGAGGCCG 596
```

```
      246 ..................................................  246

      597 GCCCCGAAAGCTGAGCAAAGAGTACTGGGACTGTCTCGAGGGCAAGAAGA  646

      247 ..............GlnProIleHisLeuTrpGlnPheLeuLysGluLeu  258
                       :::     ||||||||||:::|||:::::::::::::
      647 GCAAGCACGGGCCCAGAGGCACCCACCTGTGGGAGTTCATCCGGACACATC  696

      259 LeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGl  275
          |||:::  |||:::          :::::::::|||  |||:::::
      697 CTCATCCACCCGGAGCTCAACGAGGGCTCTATCAAGTGGGAGAACCGCCA  746

      275 GLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpG  292
          :::|||:::|||||:::       |||      ||||||:::|||||||
      747 TGAAGGCGTCTTCAAGTTCCTGCGCTCCGAGGCTGTGGCCCAACTATGGG  796

      292 IyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSer  308
          ||   :::|||:::::       ||||:|||:::|||||||||||:::
      797 GCCAAAAGAAAAAGAACAGCAACATGACCTACGACAAGCTGAGCCGGGCC  846

      309 IleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGl  325
          :::|||   ||||||||||:::   |||:::::::::   |||   :::::
      847 ATCAGGTACTACTACAAACGGGAGATCCTGGAACCGGCTGGAT...GGCCG  893

      325 nArgLeuValTyrGlnPhe 331
          :|||||||||||||:::|||
      894 GCGACTCGTCTACAAGTTT 912

Seq name: /SIDS8/gcgdata/geneseq/geneseqp/NA2000.DAT:AAC99029

Seq documentation block:
ID   AAC99029 standard; cDNA; 2952 BP.
XX
AC   AAC99029;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:257.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification, tissue typing, forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA;  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54264.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition
XX
PS   Claim 1; Page 696-697; 1479pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 2952 BP; 766 A; 645 C; 671 G; 868 T; 2 other;


Alignment scores:
            Quality:  269.50            Length:     85
              Ratio:  3.223               Gaps:      4
 Percent Similarity:  76.471   Percent Identity:  54.118

Alignment Block:
 US-09-126-945B-2 x AAC99029   ..

 Align seq 1/1  to: AAC99029  from: 1  to: 2952

      249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
          :::::|||||||||||||||||      ||||||      |||   :::
      175 CTTCAGCTGTGGCAGTTCCTGGTCACCCTTCTTGATGACCCA...GCCAA 221

      265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle...PheLys1 281
           :::::|||||   |||   :::      :::|||:::   ||||||
      222 TGCCCACTTCATTGCCTGGACAGGT.....CGAGGCATGGAGTTCAAGC 265

      281 leGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
          ::   :::      :::||||||||||   ||||||||||:::|||||||
      266 TGATAGAACCGAAGAGGTTGCTCGGCGCTGGGGCATCCAGAAGAACCGG 315

      298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 314
          ||||||||||||||||||||||||||||||||:::|||   ||||||::
      316 CCAGCCATGAACTATGACAAGCTGAGCCGCTCTCTCCGCCATTACTATGA 365

      314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
          :|||||||||:::::|||        :::::|||   ||||||:::|
      366 AAAGGGCATCATCCGAAGGCG...GCTGGAGAGCGATACGTCTACAAAT 412

      331 heVal 332
          |||||
      413 TTGTC 417

Seq_name: /SI... ...; gcgdata/geneseq/geneseqp/NA2000.DAT:AAB00768

Seq documentation block:
ID   AAB00768 standard; cDNA; 9679 BP.
XX
AC   AAB00768;
XX
DT   08-SEP-2000  (first entry)
XX
DE   Rat phosphodiesterase interacting protein, myomegalin cDNA.
XX
KW   Rat; phosphodiesterase interacting protein; myomegalin; PDE; cAMP PDE;
KW   cyclic adenosine monophosphate phosphodiesterase; antiasthmatic; asthma;
KW   antiinflammatory; antipsoriatic; dermatological; antibacterial; shock;
KW   analgesic; immunosuppressive; antiulcer; vasotropic; antiarthritic;
```

```
KW   antidiabetic; urticaria; antiallergic; antiarteriosclerotic; diagnosis;
KW   antirheumatic; treatment; inflammatory disease; psoriasis; arthritis;
KW   atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;
KW   eosinophilic granuloma; proliferative skin disease; ulcerative colitis;
KW   reperfusion injury; atopic dermatitis; diabetes insipidus;
KW   conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;
KW   arterial restenosis; ankylosing spondylitis; transplant rejection;
KW   graft versus host disease; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1746..8726
FT                   /*tag  a
FT                   /product  "Myomegalin"
FT                   /note  "This region is referred to as CDS 1 in N..  :
FT                   in the specification"
XX
PN   WO200027861-A1.
XX
PD   18-MAY-2000.
XX
PF   12-NOV-1999;   99WO-US26860.
XX
PR   12-NOV-1998;   98US-0108255
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Conti M,  Pahlke G;
XX
DR   WPI; 2000-376479/32.
DR   P-PSDB; AAY71158.
XX
PT   Polynucleotide encoding a phosphodiesterase (PDE) interacting
PT   polypeptide, useful for diagnosis and treatment of asthma, cystic
PT   fibrosis, Crohn's disease, and rheumatoid arthritis .
XX
PS   Disclosure; Fig 2; 77pp; English.
XX
CC   The present sequence is a cDNA encoding a phosphodiesterase (PDE)
CC   interacting protein, myomegalin from rat skeletal muscle cDNA library.
CC   The protein modulates the functions and properties of PDEs, specifically
CC   cAMP-PDEs, and also targets them to specific subcellular compartments.
CC   Rat myomegalin has at least four differently sized transcripts, two of
CC   which are expressed in heart (7.5 and 5.9 kb), two in skeletal muscle
CC   (7.5 and 4.3 kb) and one in testis (2.5 kb). The present sequence
CC   can be used in the diagnosis and treatment of disease conditions
CC   associated with PDE activity. The diseases include asthma, cystic
CC   fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic
CC   granuloma, psoriasis, proliferative skin diseases, endotoxic shock,
CC   septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,
CC   inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory
CC   distress syndrome, diabetes insipidus, allergic rhinitis, allergic
CC   conjunctivitis, vernal conjunctivitis, arterial restenosis,
CC   atherosclerosis, inflammatory diseases associated with irritation and
CC   pain, rheumatoid arthritis, ankylosing spondylitis, transplant
CC   rejection and graft versus host disease, disease conditions associated
CC   with hypersecretion of gastric acid, and disease conditions in which
CC   cytokines are mediators.
XX
SQ   Sequence 9679 BP; 2543 A; 2626 C; 2646 G; 1862 T; 1 other;
```

```
alignment_scores:
            Quality:  203.00              Length:     245
              Ratio:  1.664                 Gaps:       8
 Percent Similarity:  49.796   Percent Identity: 23.673

alignment_block:
 US-09-126-945B-2 x AAY00768/rev

 Align seq 1/1  to reverse of: AAY00768  from: 1  to: 9679
```

```
 99 GlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGl 115
    ::::::|||::::::::::|||||||||::      ::::::      ::
592 GAGGGACCCACGAACCAGCTGGATCTGCTCATCCGGGCCGTGGAAGCATC 543

115 uGluHisSer.....LeuGluGlnValG 128
    :  ||||||          ::||||  ::::
542 AGTTCACAGCAGTAAATGCACACTGTACGGATAAGACCATTGAAGCTGCTG 493

129 lnSerMetValValGlyGlu..... 142
    ::::::::::      |||
492 AAGCCCTGCTTCATATGGAGTCTCTACCTGCCTTGAGGATTCAAGAATCC 443

130 .....ValLeuLysAspIle.....GluThrAlaCy 148
          :::      ||::::      ||||||:::
442 GATTTATCCATGCTGCGATGAGGCCAGATGTCATTACGGAGACTGTAGT 393

149 sLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnV 195
    ::::::   :::      ::::::||||||||||   ||||||   ::::
392 GGAGGTGTCAACAGAGGAGTCTGAACCAATGGATGCCTCGCCTATCCATA 343

196 alGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMet 171
    :::::                ::||||::             ||||||
342 CTTCACCA.....GACAGCCACGAA.....CCAATG 317

172 GlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGl 188
      |||                                          ::::::::
316 AAAAACAAG.....AAAGTTGGCCG 297

189 uGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaH 205
    ::::         |||:::|||||||:::::::::
296 IAAACCAAAGACCCAGCAGTCACCAGTTTCCAATGGG..... 260

205 isLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerPro 221
                 :::          ::::|||::::::::::
259 .....TCTCCTGAGTTAGGAGTCAAGAACAAACCCAGAGAA 224

222 GlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSe 238
    |||
223 GGA..... 221

238 rGlnValAspserSerLysSerGlyGlnProIleHisLeuTrpGlnPheL 255
                      ::::|::::      :::|||||:::||||
220 ..................AAAGGAAACACAACCTATCTGTGGGAGTTTC 190

255 euLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrp 271
    ||   :::|||||          ::::::      |||:::|||:::|||
189 TTTTAGATCTGCTTCAAGATAAAAATACTTGTCCCAGGTATATTAAATGG 140

272 LeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAl 288
    :::::|||||||||||||||||||||::::   ||||||      |||::
139 ACTCAGAGAGAAAAGGGCATATTCAAGCTGGTGGACTCAAAGGCTGTCTC 90

288 gArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysL 305
    ::::|||||||||   :::|||||::::|||   ||||||||||:::   :
 89 GAAGCTCTGGGCAAAGCATAAGAACAAACCAGACATGAACTATGAAACTA 40

305 euSerArgSerIleArgGlnTyrTyrLysLysGly 336
    ::::|||:::::::|||   ||||||:::::|||
 39 TGGGGCCAGCTCTCACATACTATCAACCAAACCGCA 5
```

```
Seq_documentation_block:
ID   AAT47198 standard; cDNA; 2667 BP.
XX
AC   AAT47198;
XX
DT   06-APR-1997  (first entry)
XX
DE   Human EF2 repressor factor (ERF) cDNA.
XX
KW   EF2 repressor factor; ERF; transcriptional repressor;
KW   tumour suppressor; tumour; cancer; oncogene; gene therapy; ss.
```

```
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             123..1769
FT                   /*tag= a
FT   polyA_signal    2644..2649
FT                   /*tag= b
XX
PN   WO9639517-A1.
XX
PD   12-DEC-1996.
XX
PF   04-JUN-1996;    96WO-US10177.
XX
PR   05-JUN-1995;    95US-0469412.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Athanasiou MA,  Boal GJ,  Blair DG,  Fisher RJ,  Mavrothalassitis GJ;
PI   Sgouras D N;
XX
DR   WPI; 1997-044139/04.
DR   P-PSDB; AAW07700.
XX
PT   New DNA encoding ETS2 repressor factor useful for reducing
PT   tumourigenicity, esp. oncogene associated tumour cells
XX
PS   Claim 3; Page 63-65; 101pp; English.
XX
CC   A cDNA clone (AAI47198) codes for human ETS2 repressor factor (ERF)
CC   (AAW07700). It was isolated from a K562 cDNA library using the H1
CC   site of the ETS2 promoter as probe. A related clone (AAI47199)
CC   coding for an alternatively spliced ERF (AAW07701) was also isolated.
CC   The ERF gene, which maps to chromosome 19, q1.2-1.3, is the first
CC   member of the ets family to be identified as a transcriptional
CC   repressor in mammalian cells. It can be used to suppress or repress
CC   transcription and to elucidate transcription process and regulation.
CC   The ERF gene also has tumour suppresor activity and can be used to
CC   reduce ets-dependent tumourigenicity associated with v-mos, c-met,
CC   tpr-met, Ha-ras and qag-myb-ets oncogenes. The cDNA can be cloned
CC   for expression of the ERF polypeptide in host cells.
XX
SQ   Sequence 2667 BP; 456 A; 876 C; 825 G; 510 T; 0 other;


Alignment scores:
            Quality:  194.50            Length:      91
              Ratio:  2.860               Gaps:       4
 Percent Similarity:  74.725   Percent Identity: 47.253

Alignment block:
 US-09-126-945B-2 x AAI47198   ..

 Align seq 1/1  to: AAI47198  from: 1  to: 2667

      241 AspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGl 257
          :::|||||   :::::::::   |||:::|||||:::|||:::   ||
      177 GAGTCGTCCCCTGCCTCAAGGCAGATCCAGCTGTGGCACTTTATCCTGGA 226

      257 uLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnL 274
          |||||||   |||   :::::::|||         |||   |||   :::
      227 GCTGCTGCGGAAG...GAGGAGTACCAGGGCGTCATTGCCTGG...CAGG 270

      274 ysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeu 290
            :::   |||   |||   |||:::|||       :::|||||||||||
      271 GGGACTACGGGGAATTCGTCATCAAAGACCCTGATGAGGTGGCCCGGCTG 320

      291 TrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerAr 307
          |||||:::|||||   :::|||   |||||||||||||||||||||||||
      321 TGGGGCGTTCGCAAGTGCAAGCCCCAGATGAATTACGACAAGCTGAGCCG 370

      307 qSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleS 324
```

```
          |:::::::|||   |||||||:::|||   |||:::::::|||         :
      471 GGCCCTGCGCTATTACTATAACAAGCGGCATTCTGGCAGAAGACCAAG...G 417

      324 erGlnArgLeuValTyrGlnPhe 331
          :::::|||:::::::|||:::|||
      418 GGAAACGCTTCACCTACAAGTTC 440

>>seq_no 2 ; ss_1 ; alias score 0 ; pn:US-09-126-945B-2 ; pa:AAFI7794

seq_documentation_block:
ID   AAFI7794 standard; cDNA; 502 BP.
XX
AC   AAFI7794;
XX
DT   18-MAR-2001  (first entry)
XX
DE   Human breast cancer associated 20406 coding sequence.
XX
KW   Human; breast cancer associated gene; vaccine; diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200060076-A2.
XX
PD   12-OCT-2000.
XX
PF   15-FEB-2000; 2000WO-US05408.
XX
PR   02-APR-1999;   99US-0285480.
PR   23-JUN-1999;   99US-0339338.
PR   02-SEP-1999;   99US-0389681.
PR   03-NOV-1999;   99US-0433826.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Yuqiu J,  Dillon DC,  Mitcham JL,  Xu J,  Barlocker SL;
XX
DR   WPI; 2001-122627/13.
XX
PT   An isolated polypeptide useful for the treatment and diagnosis of
PT   tumors e.g. breast cancer comprises at least an immunogenic portion of
PT   a breast tumor protein.
XX
PS   Claim 66; Page 174; 238pp; English.
XX
CC   The present invention provides the coding sequences and some protein
CC   sequences of proteins associated with breast cancer in humans. These
CC   sequences can be used in the diagnosis and treatment of cancers,
CC   particularly breast tumours.
XX
SQ   Sequence 502 BP; 120 A; 147 C; 157 G; 78 T; 0 other;


Alignment scores:
            Quality:  182.50            Length:      93
              Ratio:  2.765               Gaps:       1
 Percent Similarity:  70.968   Percent Identity: 47.634

Alignment block:
 US-09-126-945B-2 x AAFI7794   ..

 Align seq 1/1  to: AAFI7794  from: 1  to: 502

      231 SerGluGluSerTrpThrAspSerGlu...ValAspSerSerCysSerGl 246
          |||:::|||   |||         |||         |||:::   :::
      213 AGCAAAGAGTACTGGGACTGTCTCGAGGGCAAGAAGACAAGCACGCGCC 262

      246 yGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProH 263
            :::     ||||||||||:::|||:::::::::::::|||:::   |||
      263 CAGAGGCACCCACCTGTGGGAGTTCATCCGGGACATCCTCATCCACCCGG 312

      263 isSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePhe 279
          ::          :::::::::|||   |||:::::::::|||:::|||
```

```
      413 AGCTCAACGAGGGCCTCATGAAGTGGGAGAATCGGCATGAAGGCGTCTTC 462

      280 LysIleGlnAspSerAlaGlnValAlaArgLeuArgGlyIleArgLysAs 296
          ||:::      |||      ||||||::||||||||   :::|||::
      463 AAGTTCCTGCGCTCCGAGGCTGTGGCCCAACTATGGGGCCAAAAGAAAAA 412

      296 nArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrI 313
          ::::      |||::|||:::||||||||||||||:::::|||   ||||
      413 GAACAGCAACATGACCTACGAGAAGCTGAGCCGGGCCATGAGGTACTACT 462

      314 yrLysLysGlyIleIleArgLysProAsp 322
          |||||::   |||:::::::::   |||
      463 ACAAACGGGAGATCCTGGAACGGGTGGAT 491

Seq name: ??:d:1:filtered:seq_type=seq_NA:num_as: AAV....

seq_documentation_block:
ID   AAV38865 standard; cDNA; 7752 BP.
XX
AC   AAV38865;
XX
DT   09-NOV-1998 (first entry)
XX
DE   Mouse ESX transcription factor gene.
XX
KW   ESX; epithelial-restricted with serine box; transcription factor;
KW   ETS; mouse; epithelial cancer; breast cancer; adenocarcinoma;
KW   diagnosis; therapy; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   promoter        2553..2889
FT                   /*tag= a
FT   CDS             3604..7115
FT                   /*tag= b
FT                   /note= "contains introns"
FT   exon            2900..3045
FT                   /*tag= c
FT                   /number= 1
FT   intron          3046..3606
FT                   /*tag= d
FT                   /number= 1
FT   exon            3604..3764
FT                   /*tag= e
FT                   /number= 2
FT   intron          3765..4148
FT                   /*tag= f
FT                   /number= 2
FT   exon            4149..4373
FT                   /*tag= g
FT                   /number= 3
FT   intron          4374..4503
FT                   /*tag= h
FT                   /number= 3
FT   exon            4504..4594
FT                   /*tag= i
FT                   /number= 4
FT   intron          4595..4782
FT                   /*tag= j
FT                   /number= 4
FT   exon            4783..4905
FT                   /*tag= k
FT                   /number= 5
FT   intron          4906..5052
FT                   /*tag= l
FT                   /number= 5
FT   exon            5053..5142
FT                   /*tag= m
FT                   /number= 6
FT   intron          5143..5284
FT                   /*tag= n
FT                   /number= 6
FT   exon            5285..5402
FT                   /*tag= o
FT                   /number= 7
FT   intron          5403..6255
FT                   /*tag= p
FT                   /number= 7
FT   exon            6256..6450
FT                   /*tag= q
FT                   /number= 8
FT   intron          6451..6998
FT                   /*tag= r
FT                   /number= 8
FT   exon            6999..7115
FT                   /*tag= s
FT                   /number= 9
XX
PN   WO9823782-A2.
XX
PD   04-JUN-1998.
XX
PF   26-NOV-1997;    97WO-US21865.
XX
PR   25-NOV-1997;    97US-0031504.
PR   27-NOV-1996;    96US-0031504.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Benz CC, Chang C, Scott GK;
XX
DR   WPI; 1998-322755/28.
DR   P-PSDB; AAW60680.
XX
PT   Nucleic acid encoding the ESX transcription activator -
PT   over-expressed in epithelial, especially breast, cancers, also
PT   related polypeptide(s), antibodies, vectors and transformed cells,
PT   useful for diagnosis and treatment of cancer
XX
PS   Claim 18; Page 85-90; 120pp; English.
XX
CC   This is the murine mESX gene that contains about 2.9 kb of
CC   promoter upstream of about 4.9 kb of DNA incorporating at least
CC   9 exons that specify a transcript of about 2 kb, with exons 2-9
CC   encoding the 371 amino acid mESX protein (see AAW60680). The gene
CC   is a member of the ETS family and is designated ETX for
CC   epithelial restricted with serine box. In humans, a dysfunction of
CC   the ETX gene (see AAV38354) that results in increased expression is
CC   indicative of epithelial cancer (especially breast cancer but
CC   also gastric, ovarian and lung adenocarcinoma) and of unfavourable
CC   prognosis. Antibodies, antisense nucleic acids or inactive ESX
CC   mutant proteins that inhibit activity of human ESX can be used to
CC   inhibit growth and proliferation of neoplastic cells, particularly
CC   cancer. ESX polypeptides, polynucleotides and antibodies are also
CC   used to detect ESX, particularly for diagnosis and monitoring of
CC   cancer. They can also be used to screen for specific modulators,
CC   especially potential anticancer agents. Transgenic animals are
CC   also provided.
XX
SQ   Sequence 7752 BP; 1768 A; 2050 C; 2019 G; 1914 T; 1 other;


alignment_scores:
            Quality:  180.50              Length:     478
              Ratio:  1.020                 Gaps:      20
 Percent Similarity:  46.825   Percent Identity:  23.810

alignment_block:
 US-09-126-945B-2 x AAV38865    ..

 Align seq 1/1  to: AAV38865  from: 1  to: 7752

        3 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuGluLeuEr 19
          ||||||||  ||||||||:::                           ||
     5452 TCAGCTTCCACTGGCCTCTCACAG.....................CC 5477
```

```
  19 oProAspThrValSerArgThrGlyLeuGluLysAla..............  31
     |   :::   :::::::   ||||||||||:::
5478 GCTGGAACTCATTGCATTGACGGGGCTCCATGGCATTTGTTACTGCCTCT 5527

  32 ..AlaAlaGlyAlaValGlyLeuGlu.ArgArgAspTrpSerProSerPr  47
     :::   :::::::::   ||||||  ||||||||:::
5528 TTACAGAGGCCTGCTTGGACTTAGAGAAGAAGGGAACTGAGGTCCTAGGA 5577

  47 oProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheA  64
                 |||:::::::|||   :::|||:::|||      :::
5578 GAGGCCATGGACAGAGTCCAGCCTTCCCACATTCTTCCTCTTTAACTATCC 5627

  64 spMetLeuTyrPro..............................GluAsp  70
           :::|||                                |||:::
5628 CCTGTACTTGGCCCCTGTCTTGCCTGATGGAACTTCCGATGGGGAGGAG 5677

  71 SerSerTrpAlaAlaLysAlaProGly.....................  79
     ::::::   :::::::::   ||||||
5678 GCAGCTGGTGGGTGCTCAGAGCCAGGCAGGCTGGGGTGGCTGCGACTCCA 5727

  80 .................................AlaSerSerArgGluG  85
                                      :::|||||||   :::|
5728 GGCACAGCCTGCTGAAAGAGCCTGGGCCTTGTGTGCTCCAGCTGCTCTG 5777

  85 luProPro...GluGluProGlu......GlnCysProValIleAspSer  98
     ||||||||       :::|||:::          |||   ::::::   :::
5778 AACCACCTGGAGGTCACCCCAAAGGTCTTGCCTGCCGCCTCTTCCATACA 5827

  99 GlnAlaProAla..............GlySerLeuAspLeuValProGl 110
     :::   ||||||                     ::::::   |||||||||
5828 CACAGACCTGCACGTATGCACTCCTGGCCTGCTTTCCTTCTTGTCCGAG 5877

 110 yGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValV 127
      ||||||   :::                          :::   :::::|
5878 GGGACTGTTAATTCCGGGA..................AGCTGTTTCTTGG 5909

 127 alGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIle 143
     ||   :::::::::   :::       |||   |||             :::
5910 ICCCTCAGGCTATAGGCAGCTCTCTGACCCCATGT.............GTG 5947

 144 ThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLe 160
      ::::::          ||||||||   |||
5948 CCAACTTCTCACCACCACTGGTCCCCACTGAAC................. 5980

 160 uTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAla..PheGl 176
           |||:::        ||||||      :::|||
5981 ..........CATGAG........CCCCCTCACAAAGAAGCGTGTCTCT 6011

 176 uGluLeuAla.......................................... 179
      :::|||:::
6012 GTCGCTGTCCATCTTAACCAGTTGTTTGATCCTTAACTGGTGAGAGAAIC 6061

 180 .................GlyLys:GluLeuCysAlaMetSerGluGluGlu 198
                     |||          |||   ::::::::::|||
6062 GAGCGCTCTGTGCAGTCGGCCTAGCGCATTGCATTTTGGGGCAGGAA... 6108

 191 PheArgGlnArgSerPro.............LeuGlyGlyAspValLeuHi 203
           ::::::::::|||           |||   |||      :::::
6109 ......AGGAAGCAGCCACTATAGCAATCACTAAGAGGACATTTCATATA 6159

 203 sAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrS 220
     :::|||:::      |||      |||:::          ||||||
6159 CTCCCATATGCCT...TGGCTCTTAGCCTCGTTGGGATAGGAGAGG.... 6195

 220 erProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThr 238
       ||||||   :::                 ||||||      |||
6196 ..CCAGGTCGCCTA.................GAGGAGAGGGCACCC 6242

 237 AspSer.........GluValAspSerSerCysSerGlyGlnProIleHi 250
     :::       :::   :::|||       :::   :::       ||
6223 CCAGACTGATAACTGAGGAAATCTTCCCTTGTAGCCCCAGAGGTACTCA 6272

 250 sLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyA 267
     |||||||:::|||:::::::::::::::|||:::   |||:::
6273 CCTGTGGCAGTTTATCCGAGACATCCTAATCCACCCGAGCTCAACGAAG 6322

 267 rgPheIleArgIrpIonAsnLysGluLysGlyIlePheLysIleGluAsp 283
       :::::::::|||   |||:::::::::|||:::|||||||:::
6323 GCCTCATGAACTGGGAGAACCGGCACCAGGGTGTGTTCAACTTTCTTCGC 6372

 284 SerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMe 300
     |||       |||||||:::|||||||||   :::|||:::::       ||
6373 TCAGAGGCCGTGGCCCAACTCTGGGGCCAGAAGAACGAAGAACAGCAACAT 6422

 300 tAsnTyrAspLysLeuSerArgSerIleArg 318
     |:::|||:::|||||||||||||:::|||
6423 GACCTATGAGAACTGAGCCGAGCCATGAGC 6453

Seq name: [illegible] pdf_picture_placeholder_NA:1-6453:AC:AAV20472

Seq documentation block:
ID   AAV20472 standard; DNA; 1364 BP.
XX
AC   AAV20472;
XX
DT   17-JUN-1998 (first entry)
XX
DE   Human c-spi oncogene.
XX
KW   Human; oncogene; proto-oncogene; neoplastic disease; anticancer;
KW   cancer; antisense oligonucleotide; c-spi; ds.
XX
OS   Homo sapiens.
XX
PN   US5734039-A.
XX
PD   31-MAR-1998.
XX
PF   15-SEP-1994;  94US-0306691.
XX
PR   15-SEP-1994;  94US-0306691.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Calabretta B, Skorski T;
XX
DR   WPI; 1998-229882/20.
XX
PT   Anticancer composition comprising two anti-sense oligonucleotide(s)
PT   targetting cytoplasmic and nuclear oncogene(s)
XX
PS   Claim 1; Column 143-146; 92pp; English.
XX
CC   The present sequence represents an oncogene from the present invention.
CC   The present invention describes a composition which comprises two
CC   antisense oligonucleotides. The first oligonucleotide is specific for a
CC   cytoplasmic oncogene or proto-oncogene selected from ras, raf, EGF-1,
CC   c-fms, c-fos, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-fgr and
CC   c-yes. The second oligonucleotide is specific for a nuclear oncogene or
CC   proto-oncogene selected from myc, jun, c-ets, c-fos, c-myb, B-myb,
CC   c-rel, c-var, c-ski, c-spi, cyclin-D1, PML/RAR-alpha, AML1/MTG8,
CC   E2A/pbx1 and Ret-1/AF-4. The composition is used for treating cancer.
CC   The combination of antisense oligonucleotides has synergistically
CC   enhanced ability to inhibit growth of cancer cells.
XX
SQ   Sequence 1364 BP; 279 A; 497 C; 390 G; 198 T; 0 other;


alignment scores:
            Quality:  170.00              Length:     459
              Ratio:  0.988                 Gaps:      21
 Percent Similarity:  47.911   Percent Identity: 25.905

alignment block:
 US-09-126-945B-2 x AAV20472   ..
```

```
Align_seq_1/1  to: AAV20472  from: 1  to: 1364

      20 ProAspIleValSerArgThrGlyLeuGluLysAla...AlaAlaGlyAl  35
         |||  :::|||:::       |||       :::|||  :::|||  ::
      22 CCCTGCAATGTCAAGGGAGGGGGCTCACCCAGGGCTCCTGTAGCTCAGGG  71

      35 aValGlyLeuGluArgArgAspTrpSerProSerProProAlaIhrProG  52
         :::|||||||                 ||||||:::|||          ||||
      72 GGCAGGCCTG.............AGCCCTGCACCCGCCCCACGACCGT  106

      52 luGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrPro  68
         :::::        |||                              :::|||
     107 CCAGCCCCTGACGGGCA.....................CCCCATCCT  142

      69 Glu.AspSerSerTrpAlaAluLysAlaProGlyAlaSerSerArgGluG  85
         |||    |||:::   |||            |||      :::   :::|
     143 GAGGGGCTCTGCATTGGCCCCCACCGAGGCAGGGGATCTGACCGACTCGG  192

      86 luProProGluGluProGluGlnCys.............ProValIle  96
         |||||            :::   |||               |||||:::
     193 AGCCCGGCTGGATGTTACAGGCGTGCAAAATGGAAGGGTTTCCCCTGGTC  242

      97 AspSerGluAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuTh  113
                 |||      |||   |||||||||||||
     243 ...........CCCCCTCCATCAGAAGACCTGGTGCCCTATGACACGGA  290

     113 rLeu.........GluGluHis.SerLeuGluGluValGlnSerMetVal  126
          |||           :::   ||| |||:::      :::   |||   |||
     291 TCTATAGCAACGCCAAACGCACGAGTATTACCCCTATC...TCAGCAGTG  337

     127 ValGlyGluValLeuLysAspIle.......................... 134
         :::|||:::::::::::      |||
     338 ATGGGAGAGCCATAGCGACCATTACTGGGACTTCCACCCCACCACGTG  387

     135 ..............................................GluThrA  147
                                                     :::   |
     388 CACAGCGAGTTCGAGAGCTTCGCCGAGAACAACTTCACGGAGCTCCAGAG  437

     137 laCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSer  153
         ||||:::      :::   :::::::          ||||||
     438 CGTGCAGCCCCCGCAGCTCCAGCAGCTCTACCGCCACATGGAGC......  481

     154 AsnValGlnLysTrpLeuLeuTrpThrGluHisGlnIyrArgLeuProPr  170
                          |||:::::::            :::||||
     482 ....................TGGAGCAGATGCACGTCCTCGATACCCCC  526

     170 oMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeu.......... 183
         |          :::      ||||||  ::::::
     491 ATGCTGCCACCCCATCCCAGTCTTGGCCACCAGCTCTCCTACCTCCCCG  540

     184 ..CysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGly  199
           ||||||   |||  :::          |||   ||||||
     541 GATGTGCCTCCAGTACCCATCCCTGTCCCCAGCCCAGCCCA.........  581

     200 AspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaIrpMetL;  216
                                                |||  ||:::
     582 ..................................GCTCAGATGAG  592

     216 sGluArgThrSerProGlyAlaIleHisTyr...CysAlaSerThrSerG  232
         :::|||:::|||      |||  |||:::  |||  ::::::   :
     593 GAGGAGGGCCAGCGGCAGAGCCCCCCACTGGAGTTGTCTGACGGCGAGGC  642

     232 IsGluSerTrpThrAspSerGluValAspSerSerGlySerGly...... 248
         ::  :::|||:::  :::      :::|||||  |||
     643 GGATGGCCTGGAGCCCGGCCTG.....GGCTCCTGCCTGGGGAGACAG  686

     247 ......Gln..ProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeu  261
               :::    |||:::|||:::|||||||||   :::   ||||||
     687 GCAGCAACAAGAAGATCCGCCTGTACCAGTTCCTGTTGGAC...CTGCTC  733

     261 LysProHisSerIyrGlyArgPheIleArgTrpLeuAsnLysGluLysGl 277
         :::      :::            |||   |||:::::|||:::|||||

     734 CGCAGCGGCGACATGAAGGACAGCATCTGGTGGCTGGACAAGGACAAGGG 783

     277 yIlePheLysIleGluAspSerAlaGln......ValAlaArgLeuTrpG 292
         |   |||:::::::::::::::::   :::      :::|||:::   ||||
     784 CACCTTCCAGATCTCCTTCAAGACAACAGGAAGCGCTGGCGCACCGCTGGC 833

     292 lyIleArgLys...AsnArgProAlaMetAsnIyrAspLysLeuSerArg 307
         |||||:::|||   |:|||      |||:::|||:::|||:::|||
     834 GCATCCAGAAGGGCAACCGCAAGAAGATGACCTACCAGAAGATGGCGCGC 883

     308 SerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSe 324
         :::::|||:::|||   |||   |||   :::::::|||      :::::
     884 GCGCTGCGCAACTACGGCAAGACGGGCGAGGTCAAGAAG.....GTGAA 927

     324 rGlnArgLeuValTyrGlnPhe 331
         :::::::|||:::|||||||||
     928 GAAGAACCTCACCTACCAGTTC 949

seq_name: _2188sR_ artificial_protein p_protein_gi_NA2000_DAT:AAA61647

seq_documentation_block:
ID   AAA61647 standard; cDNA; 1527 BP.
XX
AC   AAA61647;
XX
DT   23-OCT-2000  (first entry)
XX
DE   cDNA encoding rat Elf-1 polymorphic variant #3.
XX
KW   Elf-1; transcriptional regulatory factor; rat; polymorphic variant;
KW   immunological disease; allergic disease; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1527
FT                   /*tag= a
FT                   /product= "Rat Elf-1 variant #3"
FT                   /trans_exception= (pos:433..432, aa:Gly->Val)
XX
PN   JP2000135088-A.
XX
PD   16-MAY-2000.
XX
PF   30-OCT-1998;   98JP-0309595.
XX
PR   30-OCT-1998;   98JP-0309595.
XX
PA   (ASAK ) ASAHI BREWERIES LTD.
XX
DR   WPI; 2000-425802/37.
DR   P-PSDB; AAB03192.
XX
PT   A transcription regulating factor Elf-1 cDNA and its polymorphism
XX
PS   Claim 6; Page 15-17; 21pp; Japanese.
XX
CC   The invention relates to three polymorphic variants of a rat
CC   transcriptional regulatory factor Elf-1 (AAB03190-B03192) and nucleic
CC   acids encoding them (AAA61645-A61647). The invention also encompasses
CC   the use of Elf-1 for regulating the transcription of a gene in an
CC   expression vector. The cDNA encoding rat Elf-1 variant #1 (AAA61645)
CC   was isolated using PCR primers (AAA61648-A61649) based on mouse Elf-1
CC   cDNA, and the rat Elf-1 variant #2 and #3 cDNAs (AAA61646-A61647) were
CC   subsequently isolated. The Elf-1 cDNA and protein sequences are
CC   potentially useful for the prevention and treatment of immunological and
CC   allergic diseases mediated by Elf-1. The present sequence represents
CC   cDNA encoding a variant (#3) of rat Elf-1.
XX
SQ   Sequence 1527 BP; 458 A; 394 C; 345 G; 330 T; 0 other;


alignment_scores:
```

```
            Quality:  164.50              Length:      86
              Ratio:  3.046                 Gaps:       2
 Percent Similarity:  62.791   Percent Identity:  48.872

alignment block:
 US-09-126-945B-2 x AAA61647   ..

 Align seq 1/1  to: AAA61647  from: 1  to: 1527

      246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 282
          |||:::  |||:::|||||:::|||||||      |||||||
      337 GGAAACACAATTTACCTTTGGGAGTTTTTACTGGCTCTGCTTCAGGACAA 386

      282 oHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 279
            :::    ::::::|||:::|||   ::::::||||||||
      387 AGCTACCTGTCCTAAGTACATCAAATGGACACACGGAGAAAAAGGC.... 432

      279 heLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLys 295
                                                  ::::||
      433 ..................................................CACAAA 438

      296 AsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTy 812
          |||:::|||   ||||||||||:::   ::::::|||:::::|||   ||
      439 AACAAACCTGACATGAACTATGACAAGCTATCGGGCGAGCACTCAGGTACTA 488

      812 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVal1 829
          ||||:::::|||||:::   |||   :::   :::|||||||||||||
      489 TTACCAAAGAGGTATTCTTGCAAAATGGAA...GGTCAGCGCTTGGTGT 536

      829 yrGlnPhe 836
          ||||||||
      546 ATCAGTTT 553

seq name: /SIDS82/seqdata/genseq/genseqpo/NA1997.DAT:AAI47200

seq documentation block:
ID   AAI47200 standard; DNA; 2544 BP.
XX
AC   AAI47200;
XX
DT   06 APR 1997  (first entry)
XX
DE   Mouse ETS2 repressor factor (ERF) gene.
XX
KW   ETS2 repressor factor; ERF; transcriptional repressor;
KW   tumour suppressor; tumour; cancer; oncogene; gene therapy; ss.
XX
OS   Mus sp.
XX
PN   WO9639517 A1.
XX
PD   12 DEC 1996.
XX
PF   04 JUN 1996;   96WO-US10177.
XX
PR   05 JUN 1995;   95US-0469412.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Athanasiou MA,  Beal GJ,  Blair DG,  Fisher RJ,  Mavrothalassitis GJ;
PI   Sgouras D N;
XX
DR   WPI; 1997-043139/04.
DR   P-PSDB; AAW07702.
XX
PT   New DNA encoding ETS2 repressor factor - useful for reducing
PT   tumourigenicity, esp. oncogene associated tumour cells
XX
PS   Disclosure; Page 68-70; 101pp; English.
XX
CC   A murine DNA clone (AAI47200) codes for ETS2 repressor factor (ERF)
CC   (AAW07702), a member of the ets gene family identified as a
CC   transcriptional repressor in mammalian cells. Human ERF cDNA
CC   clones (see also AAI47198-99) have also been identified. ERF nucleic
CC   acids can be used to suppress or repress transcription and to
CC   elucidate transcription process and regulation. They also have
CC   tumour suppresor activity and can be used to reduce ets dependent
CC   tumorigenicity. The genes may also be cloned for expression of
CC   ERF polypeptides in host cells.
XX
SQ   Sequence 2544 BP; 448 A; 820 C; 736 G; 540 T; 0 other;


alignment scores:
            Quality:  160.00              Length:      70
              Ratio:  3.077                 Gaps:       2
 Percent Similarity:  74.286   Percent Identity:  50.000

alignment block:
 US-09-126-945B-2 x AAI47200   ..

 Align seq 1/1  to: AAI47200  from: 1  to: 2544

      241 AspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGl 257
          :::||||||   ||||||||||   |||:::|||||:::|||:::   ||
      417 GAGTCATCCCCTGCTCCAGGCAGATCCAGCTGTGGCACTTTATCCTGGA 466

      257 uLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnL 274
          |||||||   |||   :::::|||          |||   |||   :::
      467 GCTGCTTCCGAAA...GAGGAGTACCAGGGCGTCATGCTTGG...CAGG 510

      274 ysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeu 290
            :::   |||   |||   |||:::|||       :::||||||||||||
      511 GGGACTACGGGGAGTTTGTCATCAACGACCCTGATGAAGTGGCTCGCCTC 560

      291 rpGlyIleArgLysAspArgProAlaMetAsnTyrAspLysLeuSerAr 307
          ||||||:::|||||   :::|||   |||||||||||||||||||||||
      561 TGGGGGGTCCGCAAGTGCAAACCCCAGATGAACTATGACAAGCTGAGCCG 610

      307 qSerIleArg 316
          |:::::|||
      611 GGCTTTGCGG 620

seq name: /SIDS82/seqdata/genseq/genseqpo/NA1991.DAT:AAQ13187

seq documentation block:
ID   AAQ13187 standard; cDNA; 1313 BP.
XX
AC   AAQ13187;
XX
DT   30-OCT-1991  (first entry)
XX
DE   Polypeptide with DNA binding domain and transcription effecting
DE   domain DNA sequence.
XX
KW   PDL; ss.
XX
OS   Homo sapiens.
XX
PN   WO9111521-A.
XX
PD   08-AUG-1991.
XX
PF   25-JAN-1991;   91WO-US00557.
XX
PR   26-JAN-1990;   90US-0470862.
XX
PA   (LJOL-) LA JOLLA CANCER RES.
XX
PI   Maki RA,  Klemsz MJ,  McKercher SR,  Celada A;
XX
DR   WPI; 1991-252652/34.
DR   P-PSDB; AAK13444.
XX
PT   Polypeptide with DNA binding domain and transcription effecting
PT   domain - used for studying, preventing or enhancing transcription
```

```
PT   of DNA.
XX
PS   Claim 3; Fig 1B; 49pp; English.
XX
CC   The DNA encodes a polypeptide with a DNA binding domain and
CC   transcription effecting domain. The sequence was obtd. using mouse
CC   PU.1 cDNA as a probe to screen human macrophage and B-cell
CC   libraries to obtain human PU.1 cDNA. It can be used to induce
CC   tumours in animals for study by producing chimeric animals from
CC   PU.1 transfected bone marrow stem cells or PU.1 trans genic animals.
CC   The transcription of DNA which is activated by PU.1 can be prevented
CC   or reduced by altering the activity inducing phosphorylation of the
CC   protein. The N terminal of the DNA binding domain lies between
CC   amino acids 129 200 and the C-terminal end between 230 and 267.
XX
SQ   Sequence 1313 BP; 285 A; 429 C; 367 G; 232 T; 0 other;


alignment_scores:
            Quality:  155.00              Length:     308
              Ratio:  0.957                 Gaps:      21
 Percent Similarity:  52.597   Percent Identity:  25.974

alignment_block:
 US-09-126-945B-2 x AAQ13187   ..

 Align Seq 1/1  to: AAQ13187  from: 1  to: 1313

       86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
          |||   :::::::|||:::                 |||:::   |||::
       21 CCTGCGTCTGACCCACGA.............CCGTCCAGTCCCCCGAC 56

      102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHis.Ser 118
          :|||      ||||||      |||:::   |||:::   ||| :::
       56 GGGG......CACCTGGTCCTGAGGGGATCCGCCTTGATCCCCACCGAA 99

      119 LeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspIleGl 135
          ::::::   ::::::   ::::::   :::|||:::
      100 GCAGGGATCTGACCAACCTGGAGCTCAGCTGGATGTTACAG........ 141

      135 qThrAlaCysLysLeu.....LeuAsnIleThrAlaAspProMet...A 149
          ||||||||:::       :::::::::||||||   |||      |
      142 ....GCGTGCAAAATGGAACGGTTTCCCTCATCCGCCCCTCCATCGGATG 187

      149 sp.Trp..............SerProSerAsnValGln..... 156
          || |||               |||   :::|||||||||
      188 ACTTGGTTACTTACGATTCAGAGCTATACCAACGTCCAATGCATGACTAC 237

      157 .......LysTrpLeuLeu.TrpThrGlu..................... 168
                ::|||         |||   :::
      238 TACTCCTTCGTGGGCAGCGATGGAGAAAGCCATAGCGATCACTACTGCGA 287

      164 ..................HisGluTyrArg.........LeuPro..... 179
                             |||   |||         ||||||
      288 TTTCTCCGCACACCATGTCCACAACAACGAGTTTGAGAACTTCCCTGACA 337

      170 .ProMetGlyLysAlaPheGluGluAlaGlyLysGluLeuCysAlaM 186
           ||:::   :::|||   :::::   |||             :::|||
      338 ACCACTTCACAGAGCTGCAGAGCGTGCAGCCCCCGCAGCTACAGCAGCTC 387

      186 etSerGluGlu.GlnPheArgGlnArgSerProLeuGlyGlyAspValIle 202
          ::|||   ::: ::::::::::|||         |||       :::::
      388 TATGGCCACATGGAGCTGGAACAGATGCACCTCCTCGATACTCCCATGGT 437

      202 uHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgT 219
          :   |||  :::      ::::::::::::::|||   :::
      438 GCCACCCCACACCGGCCTCAGTCACCAGGTTCCTACATGCCCCGATGT 487

      220 hrSerPro           ...GlyAlaIleHisTyrCysAlaSerThrCys 237
                |||          :::      |||     :::|||   :::
      488 GCTTCCCTTATCAAACCTTGTCCCCAGCCCACCAGCAGAGCTCAGATGAG 537

      238 GluGln.....................SerTrpThrAspSerGluVa 249
          ||||||                        :::   :::|||::|||::
      538 GAGGAGGCTCAGAGCAGAGCCCTCCCCTGGAGGTGTCTGATGGAGAAGC 587

      240 lAsp.................................SerSerCysSerG 246
          :|||                                ::::::   |||
      588 TGATGGCTTGGAGCCTGGGCCAGGTCTTCTGCACGGGGAGACAGGCAGCA 637

      246 lyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPro 262
            :::   |||:::|||:::|||||||||   :::   |||||||::
      638 AGAAAAAGATTGGCCTGTACCAGTTCCTGCTGGACTTCTGCTGCAGC 684

      263 HisSerTyrGlyArgPheIleArgTrpLeuAspLysGluLysGlyIlePh 279
            :::           |||   |||:::::::|||::|||||   ||
      685 GCCGACATGAAGGACAGCATCTGGTGGTGGACAAGGACAAAGGTACCTT 734

      279 eLysIleGluAspSerAlaGln......ValAlaArgLeuTrpGlyIleA 294
          |:::::::::::::::   :::      :::|||:::   |||||||||:
      735 CCAGTTCTCGTCCAAGCACAAGGAGGTCGTCTGCGCACCGCTGGGGCATCC 784

      294 rgLys...AspArgProAlaMetAsnTyrAspLysLeuSerArgSerIle 309
          ::|||   ||||||      |||:::|||:::|||:::::|||:::::
      785 AGAAGGGCAACCGCAAGAAGATGAC TACGAGAAGATGGCGCGCGCGCTG 834

      310 ArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnAr 326
          |||:::|||   |||   |||   :::::::|||      ::::::::::
      835 CGCAACTACGGCAAGACAGGCGAGGTGAAGAAA......GTCAAGAAGAA 878

      326 gLeuValTyrGlnPhe 331
          :|||:::|||||||||
      879 GCTCACCTACCAGTTC 894

Seq name  ...:US-09-126-945B-2 x AAQ13187

seq_documentation_block:
ID   AAF92329 standard; cDNA; 654 BP.
XX
AC   AAF92329;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Bovine mammary tissue derived cDNA #42.
XX
KW   Bovine; mammary gland; cancer; tumour; angiogenesis; ss.
XX
OS   Bos taurus.
XX
PN   WO200114553-A1.
XX
PD   01-MAR-2001.
XX
PF   23-AUG-2000; 2000WO-NZ00166.
XX
PR   23-AUG-1999;   99US-0150330.
XX
PA   (GENE ) GENESIS RES & DEV CORP LTD.
PA   (NZPA ) NEW ZEALAND PASTORAL AGRIC RES INST LTD.
XX
PI   Havukkala I.J.  Gleen M.  Grigor MR.  Molenaar A.J;
XX
DR   WPI; 2001-226619/23.
XX
PT   New polypeptides and polynucleotides encoding the polypeptides, which
PT   are expressed in bovine mammary gland tissue, useful for stimulating
PT   mammary gland growth or function, or inducing differentiation of milk
PT   producing cells -
XX
PS   Claim 1; Page 53; 97pp; English.
XX
CC   The present invention relates to proteins derived from bovine
CC   mammary gland cells. The invention is useful for stimulating
CC   bovine mammary gland cell growth and function, inhibiting the
CC   growth of various mammary gland cancer cells, inhibiting
CC   angiogenesis and vascularization of tumours, or modulating
```

```
CC   the growth of blood vessels in a mammal.
XX
SQ   Sequence 654 BP; 151 A; 187 C; 157 G; 159 T; 0 other;


alignment scores:
            Quality:  154.00              Length:     151
              Ratio:  1.855                 Gaps:       2
 Percent Similarity:  54.967   Percent Identity: 27.815

alignment block:
 US-09-126-945B-2 x AAF92829   ..

 Align seq 1/1  to: AAF92829  from: 1  to: 654

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
          :::||||||||||                 ::::::   |||      ||||::
      200 CAGACAGCTTGCCACTCCTACTGGACATCCGTCCACCCCGAATACTGGAC 249

      151 rProSerAsnValGluTyrTrpLeuLeuThrThrGluHisGluTyrArgL 169
          :      :::|||   :::|||||||   :::          |||||||:::|
      250 GAAGCGCCACGTCTGGGAATGGCTCCAGTTCTGCTGTGACCAGTACAACGC 299

      168 euProProMetGlyLysAlaPhe.....GlnGluLeuAlaGlyLysGlu 182
          ||                 :::|||          ||||||||||||   :::
      300 TGGACGCCAACTGCATCTCTTTCTGCCATTTCAACATCAGTGGCCTGCAG 349

      183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199
          |||||||:::|||:::::::|||:::|||   :::   :::   :::   ||
      350 CTGTGCGGCATGACACAGGAGGAGTTCATGGAGCCGGCCGGCGTCTGTGG 399

      199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTyp..... 214
          |:::   |||:::      |||:::      :::|||   ::::::
      400 GGAGTATCTGTACTTTATCCTCCAGAGCATCCGCTCACAAGGTTACTCCT 449

      215 .....MetLysGluArgThrSerProGlyAlaIle 224
                 |||          :::      |||
      450 TTTTTAATGATCCTGATGAGACCAAGGCCACCCTCAAAGACTATGCTGAT 499

      225 HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluValAs 241
              |||   :::||||||       :::       |||   :::
      500 TCCAGTTGCTTGAAAACAAGTGGCATCAAAAGTCAAGACTGTCACAGTCA 549

      241 pSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGlu 258
          |||      |||   |||       |||||||||:::|||:::::::::|
      550 TAGTCGAACAAGCTCCAGAGTTCTATCTATGCGAATTGTGCGCAGATC 599

      258 euLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 274
          ||||||||:::|||:::::::          :::::::::||   ::::||
      600 TGCTTCTATCTCCTGAGCAAAACTGCGGCATTCTGCAATGGCAAGATAAG 649

      275 Glu 275
          |||
      650 GAA 652

Seq name: ???:??:? dir:/???/seq_???? + pn:NA2000:DAT:AAC21875

seq documentation block:
ID   AAC21875 standard; cDNA; 418 BP.
XX
AC   AAC21875;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 25950.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA library;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GESE ) GENSET.
XX
PI   Dumas Milne Edwards J;  Duclert A;  Giordano J;
XX
DR   WPI; 2000-520381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS   Claim 1; SEQ ID 25950; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No cDNA has yet been exclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 40 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 418 BP; 77 A; 97 C; 80 G; 64 T; 0 other;


alignment scores:
            Quality:  145.50              Length:     72
              Ratio:  2.910                 Gaps:        4
 Percent Similarity:  69.444   Percent Identity: 51.389

alignment block:
 US-09-126-945B-2 x AAC21875   ..

 Align seq 1/1  to: AAC21875  from: 1  to: 418

      236 ThrAspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTr 252
          |||:::|||:::::::   :::      |||||||||   |||:::||||
      107 ACAAGTAGCCGCCTTCCAAATCCAGGCAGTGGCCAG...ATCCAGCTTTG 153

      252 pGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheI 269
          ||||||||||   |||   ||||||:::      |||   :::      |
      154 GCAGTTCCTCCTGGAG...CTCCTGTCGGACAGCTCCAACTCCAGCTGCA 200

      269 leArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAla 285
          ||   |||   :::       :::|||   ||||||:::   |||
      201 TCAGCTGG...GAAGGCACCAACGGGGAGTTCAAGATGACGGATCCCGAC 247

      286 GluValAlaArgLeuTrpGlyIleArgLysAspArgProAlaMetAsnTy 302
          :::|||||||||   ||||||   ||||||:::::|||   ||||||||
      248 GAGGTGGCGCGGCTCTGGGGAGAGCGGAAGAGCAAACCCAACAGAACTA 297

      302 rAspLysLeuSerArg 307
          |||||||||||||||
      298 CGATAAGCTCAGCGC 313

Seq name: ???:??:? dir:/???/seq_???? + pn:NA2000:DAT:AAA43709

seq_documentation block:
ID   AAA43709 standard; cDNA; 445 BP.
XX
AC   AAA43709;
XX
```

```
DT   21-AUG-2000  (first entry)
XX
DE   Human secreted expressed sequence tag SEQ ID NO:294.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW   vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW   tumour; infection; depression; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200021991-A1.
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;   99WO-US24206.
XX
PR   15-OCT-1998;   98US-0104436.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR;
XX
DR   WPI; 2000-317938/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (sESTs), useful for treating various disorders
PT   such as inflammatory, infectious, and cellular proliferative disorders
XX
PS   Claim 1; Page 276; 804pp; English.
XX
CC   AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC   sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC   tissue sources. The sESTs can have a range of activities depending on
CC   the tissues they were isolated from. The activities include:
CC   chemotactic; proliferative; immunomodulatory; haematopoietic;
CC   chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC   cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC   antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC   nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used for gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC   disease), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 445 BP; 120 A; 124 C; 92 G; 109 T; 0 other;


alignment_scores:
            Quality:  145.000              Length:     130
              Ratio:  1.986                  Gaps:       7
  Percent Similarity:  56.154   Percent Identity:  44.077

alignment_block:
 US-09-126-945B-2 x AAA43709    ..

 Align seq 1/1  to: AAA43709  from: 1  to: 445

      138 CysLysLeuLeuAsnIleThrAla...AspProMetAspTrpSerProSe 153
          |||  :::|||:::   |||:::  :::|.|:::  |||||  ::
       14 TGCCTCATGCTCCTCCAACAGGCTTGCAGCCAATTTACTGGAGCAGGGA 63

      154 rAsnValGlnLysTrpLeuLeuTrpThrGluHisGluTyrArgLeuProP 170
          :::|||  :::|||||  |||:::|||||||||  |||  |
       64 TGACGTAGCCCAGTGGCTCAACTGGGCTGAAAATGAGTTTCTTTAAGGC 113

      171 roMetGlyLysAlaPheGlnGlnLeuAlaGlyLysGluLeuCysAlaMet 186
          ||:::  :::        |||:::  ||||||  |||     :::
      114 CAATTGACAGCAACACGTTTGAAATGAATGGCAAAGCTCTCCTGCTGCTG 163

      187 SerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuHi 203
          :::::|||:::|||||  ||||||||  :::|||||||||||:
      164 ACCAAAGAGGACTTTCGCTATCGATCTCCTCATTCAGGTGATGTGCTCTA 213

      203 s.....AlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysG 217
          :     |||:::              :::|||:
      214 TGAACTCCTTCAGCATATT.....CTGAAGC 239

      217 luArgThrSerPro.GlyAlaIleHisTyrCysAlaSerThrSerGlnL 233
          :|||  :::||| |||  :::  :::  :::|||||  |||
      240 AGAGG...AAACCTCGGATTCTTTTTTCACCATTCTTCCACCCTGGAAAC 286

      233 uSerTrpThrAspSerGluValAspSerSerCys..... 244
          :|||  |||:::     |||  |||
      287 TCTATACACACACAGCCGG...AGGTCATACTGCATCAGAACCATGAAGA 333

      245 .....SerGlyGlnProIleHisLeuTrp 252
                     |||  |||  |||  |||
      334 AGATAACTCTGTCCAGAGGACCCCCAGGCCATCCGTGG 371

seq_documentation_block:
ID   AAF16096 standard; cDNA; 1466 BP.
XX
AC   AAF16096;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human prostate cancer antigen nucleotide sequence SEQ ID NO:531.
XX
KW   Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW   neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW   vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW   antibacterial; gene therapy; neural; immune; reproductive; renal;
KW   gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW   wound; infectious disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055174-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000;  2000WO-US05988.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587513/55.
DR   P-PSDB; AAB56893.
```

```
XX
PT   Prostate cancer associated gene sequences, referred to as prostate
PT   cancer antigens, useful for treatment, prevention, and diagnosis of
PT   disorders such as prostate cancer.
XX
PS   Claim 1; Page 993; 2338pp; English.
XX
CC   AAF15566 to AAF16505 encode the human prostate cancer associated
CC   proteins, called prostate cancer antigens, given in AAB56663 to AAB57402.
CC   The prostate cancer antigens can have neuroprotective, cytostatic,
CC   cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC   nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC   and can be used in gene therapy. The prostate cancer antigen
CC   polynucleotides may be used for detection of prostate cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The prostate cancer antigens may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC   AAB57403 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1466 BP; 418 A; 453 C; 426 G; 269 T; 0 other;


Alignment scores:
            Quality:  109.00              Length:     968
              Ratio:  0.712               Gaps:        22
 Percent Similarity:  41.576   Percent Identity:  22.554

Alignment block:
 US-09-126-945B-2 x AAF16096  ..

 Align seq 1/1  to: AAF16096  from: 1  to: 1466

        3 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr  19
          |||   :::|||:::|||     :::|||     :::|||||||||||
      458 AGCCGGGCCCTGCCTTG...TGCTCAGCTCTCGAGAGCTGCTGCTGCC  499

       19 o..............ProAspThrValSerArgThrGly.........  27
          |            ||||||   |||     ::::::
      400 GGGTGACCTGATCCAACCTGATAAGGTGCCATCTTCAGCTACCACTGCAA 449

       28 ..........................LeuGluLysAlaAlaAlaGly  44
                                   :::   ||||||   |||
      450 GGCCCTGAGGGCAACAGCAGCACGGCACTGCCCACCCGGCTGCTCAIGGC 499

       45 AlaVal.................GlyLeuGluArgArgAspTrpSerPr  45
             |||                |||   ::|||              ||
      500 CTGGTGCCAGCTGGCAGTCCTCCCGGCACTTCGAGGCCACTGAGCCACCC 549

       45 oSerProProAlaThrPro.GlnGlnGlyLeuSerAlaPheTyrLeuSer  61
          ||||   |||:::   |||    ||||||
      550 TTCCAGCCCCAGCCCACCATGGACAGGG..................... 578

       62 IyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaPr  78
                     ||||||   :::|||||||   :::::::::::||||
      579 ............TATCCAGCTTCCTCCCCAACCTCGTCCTCTGCCCC 614

       78 oGlyAlaSerSerArgGluGluProProGluGluProGluGlnCysProV  95
          |                        |||      |||::::::   ||||
      614 T..................GAGCCAGTGACGCCCAAGGACATGCCTG 642

       95 alIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGly 111
          ||   :::
      643 TTACCCAG..................................... 658

      112 LeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValVal 128
            :::|||   :::|||   |||      |||::::::|||:::|||
      659 ...GTCCTGTACCAGCACTAGCTG.....GTCAAGGGCATGACAGTGCT 691

      128 yGluValLeuLysAspIleGlu......ThrAlaCysLysLeuLeuAsnI 144
           |||::::::   :::|||:::      |||:::   :::|||||||:::

      692 GGAGGCCCTCTGGAGATCCAGGCCATCACTGGCAGCAGGTGCTCTCCA 741

      144 leThrAlaAspProMetAsp.............TrpSerProSerAsn 154
          :::::      |||                   |||:::|||:::::
      742 TGGTGCCAGGGCCCGCCAGGCCACCAGGCTCATGCTGGGACCCAACCCAG 791

      155 ...ValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArg..... 167
             ::::::   ||||||||||      :::|||      |||
      792 TGCACAAGGACTTGGCTGCTG......AGCCACACACCCAGGAGAAGGTG 845

      168 ..........LeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyL 181
                   ||||||                              :
      846 GATAAGTGGGGCTACCA.................................A 862

      181 ysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeu 197
          ::      |||   :::::||||||                  ||||||
      863 GGGCTTCCTGCAGGCTAGGGGAGGAGCCA.............CCCCCGCTT 890

      198 .....................GlyGlyAspValLeuHisAlaHisLeuAs 207
                                |||   ::::::          |||   ::
      891 CCCTATTGTGACCAGGCCTATGGGAGGAGCTGTCCATACCCCACCGTGA 940

      207 pIleTrpLys:SerAlaAlaTrpMetLysGluArgThrSerProGlyAlaI 224
          :   |||          |||||||::
      941 GACCTGG.........GCCTGGCTC........................... 956

      224 leHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluVal 240
                                                 |||::::::
      957 ..........................................CAAGGACA 965

      241 AspSerSerCysSerGlyGlnPro.....IleHisLeuTrpGluPheLeuLy 256
          |||::::::      |||   |||   :::          :::::|||
      966 GACACCGCCTGGCCTGGTGCTCCAGGGGTGAAGCAGGCCAGAATCCTGGG 1015

      256 sGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuA 273
          |||||||:||||         ::::::   :::          |||   :
     1016 GGAGCTGCTCCTGATTTGAGCTGCATTCAGGAAGTGCGGGACAIGGTAGG 1065

      273 snLysGluLysGlyIlePheIysIleGluAspSerAlaGluValAlaArg 289
          ::::::::|||:::::                ::::::      ::::::
     1066 GGAGGCAAAAAGCCTTG.............GGCACTACCCTCCCTGTGGAG 1104

      290 LeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSe 306
          |||:::|||:::|||:::                          |||::
     1104 CTGTTCGGTGTCGGTCGAGCTAGC.............CACACCCTGAC 1148

      306 rArgSerIleArgGluTyrTyrLysLysGlyIleIleArgLysProAspI 323
          :      |||          :::||||||||:::      ::::||   :
     1149 ACCATGTTCAAGGGTACCGGAAGAGAAGGGTGTCTGCCCCCAACCTCCCC 1198

      323 le 323
          ::
     1199 TG 1200

Seq name: /S1=5/1: US-09-126-945B-2/c Seq Seq#: 1; pCT/US99/16711 AAZ24399

Seq documentation block:
ID   AAZ24399 standard; cDNA; 1467 BP.
XX
AC   AAZ24399;
XX
DT   14-FEB-2000  (first entry)
XX
DE   Human bladder tumour cDNA library derived EST 11.
XX
KW   Expressed sequence tag; human; bladder; tumour; cancer; cytostatic;
KW   treatment; gene therapy; EST; ss.
XX
OS   Homo sapiens.
XX
PN   DE19818619-A1.
XX
PD   28-OCT-1999.
```

```
XX
PF   21-APR-1998;   98DE-1018619.
XX
PR   21-APR-1998;   98DE-1018619.
XX
PA   (META ) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A;  Specht T;  Hinzmann B;  Schmitt A;  Pilarsky C;  Dahl E;
XX
DR   WPI; 1999-612028/53.
XX
PT   New nucleic acid sequences expressed in bladder tumor tissue, and
PT   derived polypeptides, for treatment of bladder tumor and identification
PT   of therapeutic agents
XX
PS   Claim 8; Page 68-69; 132pp; German.
XX
CC   This invention describes novel polypeptide fragments (I) and the
CC   polynucleotides (II) that encode them that are highly expressed in a
CC   human bladder tumour and which have cytostatic activity. (II) are used
CC   for recombinant expression of (I) and to isolate complete genes. (I) are
CC   used to identify agents suitable for treatment of bladder cancer, to
CC   directly treat this form of cancer (including expression from gene
CC   therapy vectors) or are used in a preparation for cancer treatment. (I)
CC   is also used for the generation of specific antibodies. (II) are
CC   identified by assembling ESTs (expressed sequence tags) from a
CC   particular tissue type before comparison of expression patterns. This
CC   allows a significantly longer fragment of the gene to be revealed, and
CC   therefore reduces the number of failures associated with the fact that
CC   ESTs from different libraries may represent different parts of the same
CC   unknown gene, distorting the estimated frequency of occurrence in a
CC   particular tissue. AA224268 Z43309 represent expressed sequence tag (EST)
CC   fragments isolated from a human bladder tumour cDNA library which encode
CC   the proteins represented in AAY66143-Y66198.
XX
SQ   Sequence 1467 BP; 307 A; 464 C; 432 G; 265 T; 0 other;


alignment_scores:
            Quality:  109.00              Length:     368
              Ratio:  0.712                 Gaps:      22
 Percent Similarity:  41.576   Percent Identity: 22.554

alignment_block:
 US-09-126-945B-2 x AAZ24399   ..

 Align seq 1/1  to: AAZ24399  from: 1  to: 1467

        3 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
          |||   :::|||:::|||      :::|||      :::|||||||||||
      371 AGCCGGGCCCCTGCCTTG...TGCCTCAGCTCTCGAGAGCTGCTGCTGCC 417

       19 o..............ProAspThrValSerArgThrGly.......... 27
          |             ||||||   |||      ::::::
      418 GGGTGACCTGATCCAACCTGATAAGGTGCCATCTTCAGCTACCACTGCAA 467

       28 ............................LeuGluLysAlaAlaAlaGly 44
                                      :::   ||||||   |||
      468 GGCCCTGAGGGCAACAGCAGCACGGCACTGCCCACCCGGCTGCTGATGCC 517

       45 AlaVal.................GlyLeuGluArgArgAspTrpSerPr 45
             |||                |||   :::|||           ||
      518 CTGGTGCCAGCTGGGAGTCCTCCCGGCACTTCGAGGCCACTGAGCCACCC 567

       45 oSerProProAlaThrPro.GluGlnGlyLeuSerAlaPheTyrLeuSer 61
          |||   |||:::   |||    ||||||
      568 TTCCAGCCCCAGCCCACCATGGACAGGGG..................... 596

       62 TyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaPr 78
                   ||||||   :::||||||   :::::::::::| |
      597 ..............TATCCAGCTTCCTCCTCAACCTCGTCCTCTGCCCC 641

       78 oGlyAlaSerSerArgGluGluProProGluGluProGluGlnCysProV 95
```

```
          |                        |||      |||:::::   ||||
      642 T.....................GAGCCAGTGACGCCCAAGGACATGCCTG 680

       95 alIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGly 111
          ||   :::
      681 TTACCCAG..................................... 688

      112 LeuThrLeuGluGluHisSerLeuGlnGlnValGlnSerMetValValGl 128
             :::|||   :::|||   |:|      |||:::::::|:::|||
      689 ...GTCCTGTACCAGCACTAGCTG.....GTCAAGGGCATGACAGTGCT 709

      128 yGluValLeuLysAspIleGln.....ThrAlaCysLysLeuLeuAsnI 143
           |||:::::   :::|||:::      |||:::   :::|||||:::::
      710 GGAGGCCGTCTTGGAGATCCAGGCCATCACTGGCAGCAGGCTGCTCTCCA 759

      143 leThrAlaAspProMetAsp.............TrpSerProSerAsn 154
          :::::     |||                   |||:::|||:::::
      760 TGGTGCCAGGGCCCGCCAGCCACCAGGCTCATGCTGGGACCCAACCCAG 809

      155 ...ValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArg..... 167
             :::::   ||||||||      :::|||     |||
      810 TGCACAAGGACTTGGCTGCTG.....AGCCACACACCCAGGAGAAGGTG 858

      168 .........IeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyL 181
                  ||||||                              :
      854 GATAAGTGGGCTACCA................................A 870

      181 ysGluLeuCysAlaMetSerGlnGluGlnPheArgGlnArgSerProLeu 197
          ::     |||   :::::|||||                  ||||||
      871 GGGCTTCCTGCAGGCTAGGGGAGGAGCCA...........CCCCCGCTT 908

      198 .....................GlyGlyAspValLeuHisAlaHisLeuAs 207
                                |||   :::::         |||   ::
      909 CCCTATTGTGACCAGGCCTATGGGGAGGAGCTGTCCATACGCCACCGTGA 958

      207 pIleTrpLysSerAlaAlaTrpMetLysGlnArgThrSerProGlyAlaI 224
          :   |||          |||||:::
      959 GACCTGG.........GCCTGGCTC..................... 974

      224 leHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluVal 240
                                                |||:::::::
      975 ..........................................TCAAGGACA 983

      241 AspSerSerCysSerGlyGlnPro...IleHisLeuTrpGlnPheLeuLy 256
          |||:::::      |||   |||   :::         :::::|||
      984 GACACCGCCTGGCCTGGTGCTCCAGGGGTGAAGCAGGCCAGAATCCTGGG 1033

      256 sGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuA 273
          ||||||||||||      :::::   :::         |||   :
     1034 GGAGCTGCTCCTGGTTGAGCTGCATTCAGGAAGTGCGGGACATGGTAGG 1083

      273 snLysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArg 289
          :::::::::|||::::::                :::::      :::::
     1084 GGAGGCAAAAAGCCTTG.............GGCACTACCCTCCCTGTGGAG 1121

      290 LeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSe 306
          |||:::|||:::|||:::                              |||::
     1122 CTGTTCGGTGTCCGTCGAGCTAGC..............CACACCCTGAC 1156

      306 rArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspI 323
          :        |||         :::|||||:::      :::|||   :
     1157 ACCATGTTCAAGGGTACCGGAAGAAGGGTGTCTGCCCCCAACCTCCCC 1206

      323 le 323
          ::
     1207 TG 1208

seq_name: /SIDS_Publication_Data/Publication_Seq/Publication_Seq_DNA1997.DAT:AAT96643

seq_documentation_block:
ID   AAT96643 standard; cDNA; 1733 BP.
XX
AC   AAT96643;
```

```
XX
DT   27 APR 1998  (first entry)
XX
DE   Human TULP2 cDNA.
XX
KW   TULP; TULP2; sensory neuron; neurosensory defect;
KW   cochlear degeneration; hearing loss; deafness; retinal dystrophy;
KW   retinitis pigmentosa; combined rod cone dystrophy; obesity; human;
KW   animal model; transgenic animal; therapy; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             94..1656
FT                   /*tag   a
FT   exon            1..92
FT                   /*tag   b
FT                   /number  1
FT   exon            93..125
FT                   /*tag   c
FT                   /number  2
FT   exon            126..177
FT                   /*tag   d
FT                   /number  3
FT   exon            178..302
FT                   /*tag   e
FT                   /number  4
FT   exon            303..441
FT                   /*tag   f
FT                   /number  5
FT   exon            442..738
FT                   /*tag   g
FT                   /number  6
FT   exon            739..1041
FT                   /*tag   h
FT                   /number  7
FT   exon            1042..1151
FT                   /*tag   i
FT                   /number  8
FT   exon            1152..1269
FT                   /*tag   j
FT                   /number  9
FT   exon            1270..1368
FT                   /*tag   k
FT                   /number  10
FT   exon            1369..1540
FT                   /*tag   l
FT                   /number  11
FT   exon            1541..1734
FT                   /*tag   m
FT                   /number  12
XX
PN   WO9738004 A1.
XX
PD   16 OCT 1997.
XX
PF   10 APR 1997;    97WO US05903.
XX
PR   17 SEP 1996;    96US 0714991.
PR   10 APR 1996;    96US 0630592.
PR   22 AUG 1996;    96US 0701480.
PR   04 SEP 1996;    96US 0706292.
XX
PA   (JACK ) JACKSON LAB.
PA   (SEQU ) SEQUANA THERAPEUTICS INC.
XX
PI   Naggert J;  Nishina P;  Noben Trauth K;  North M;
XX
DR   WPI; 1997-512642/47.
DR   P-PSDB; AAW86491.
XX
PT   Mammalian TULP protein  used for detecting pre-disposition to
PT   neuro sensory defects
XX
PS   Claim 5; Page 57; 89pp; English.
XX
CC   This sequence comprises human TULP2 cDNA that includes a coding
CC   region for a 520 amino acid TULP2 protein (see AAW86491).  It was
CC   isolated from a human retinal cDNA library.  Expression of TULP2
CC   is restricted to the retina and testis.  Retinal expression in
CC   adult tissue is relatively low.  The TULP2 locus is associated with
CC   a genetic predisposition to combined rod cone dystrophy (CRD).  It is
CC   tightly linked to framework marker WI-9028 on chromosome 19q.
CC   TULP2 is a member of the mammalian TULP gene family associated with
CC   various defects in sensory neurons such as cochlear defects,
CC   retinitis pigmentosa and CRD.  TULP family nucleic acids (see
CC   AAP96636-87) can be used: to identify related sequences; for
CC   production of encoded protein (see AAW86490-97); to study associated
CC   physiological pathways in vivo and in vitro; to breed transgenic or
CC   knockout animal models of disease; and to modulate (e.g. using
CC   antisense sequences) gene activity for diagnostic, prophylactic and
CC   therapeutic purposes, such as treatment of neurosensory defects.
CC   The DNA is further used as a diagnostic for genetic predisposition
CC   to the specific genetically linked defect.
XX
SQ   Sequence 1734 bp; 449 A; 480 C; 455 G; 349 T; 0 other;


alignment_scores:
           Quality:  108.50              Length:     514
             Ratio:  0.764                 Gaps:      17
 Percent Similarity:  45.223   Percent Identity: 22.611

alignment block:
 US-09-126-945B-2 x AAW96643  ..

 Align seq 1/1  to: AAW96643  from: 1  to: 1734

       31 AlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPr 47
          :::|||:::||:|||            :::|||   |||:::|||   ||
      274 ICTGCGGGAGGAGCGCCTTTTAGGTGACAGAGGCCTTGGGAACCCTTICC 322

       47 oProAlaThrProGluGluGlyLeuSerAlaPheTyrLeuSerTyrPheA 64
          ||||            :::|||   |||||||:::
      323 TCCGGAAGAAAGTGTCAGAGGCACATCTGCCCTC.............. 356

       64 spMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGly... 79
                                    |||        :::::|||||
      357 ........................TGGCATCCACAGTGCCCTGGGCAC 380

       80 ..................AlaSerSerArgGluGluProProGluGluPr 90
                           :::::|||   :::||||||:::
      381 CGTGAGCTGTGGTGGAGACCACAGGGCGAGCGCGGCTCCCGACACCGC 430

       90 oGluGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspL 107
          ::::::      |||   :::|||:::::|||:::      |||   |
      431 GGACAGAAGCAGTGTTCAGGAATCTCGGTTCTCCAGTCCCCTTTCTTATCC 480

      107 euValProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGl 123
          || :::|||   :::          |||          |||||:::|||
      481 TGGCTCCCAGACAATTCCGATGCAGAA........TTGGAGGAAGTC.. 519

      123 nSerMetValValGlyGluValLeuLysAspIleGluThrAlaCysLysL 140
                                  :::::|||:::::::
      520 .........................TCCGTGGAGAATGGT........ 544

      140 euLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGln 156
                :::::::::      ||::::      ||||||              :::
      545 .....TCCGTCTCTCCCCCACCTTTTAAACAGTCTCCGAGAATCCGACGC 579

      157 LysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLy 173
          |||          |||      ||||||            |||      ::
      580 AAGGGT......TGGCAAGCCCACCAA.........CGACCTGGGACCCG 614

      174 sAlaPheGluGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluG 190
          :|||      |||      ::::::::::::          ::::::
```

```
      615 IGCACAGGGTGACAGTGACTCCCAGGATATGGCAGATGCACACAAGTCAC 664

      193 InPheArgGlnArgSerPro...LeuGlyGlyAspValLeuHisAlaHis 235
                  :::|||   :::   ||||||   ::::::   :::
      665 CCAATATGGGACCAAACCCTGGAATGGATGGTGACTGTTATATGAAAAC 714

      206 LeuAspIleTrpLysSerAlaAlaTrpMetLysGluArg..... 236
          |||   :::   ||:::            |||:::|||
      715 TTGGCCTTCCAAAAGGAAGAAGACTTGGAAAAGAAGACAGAGGCCTCTGA 764

      219 .....IhrSerProGlyAlaIleHis.....TyrCysAlaS 229
                |||:::   :::|||   |||              :
      765 GTCTACAGGGACGAACTCCTCAGCAGCACACAACGAAGAGTTGTCCAAGG 814

      229 orThrSerGluGluSerTrpThrAspSerGlu..... 259
          ::   :::   |||:::   ||||||||||:::
      815 CCCTGAAAGGCCAGGCTGGCACGGCACAGCGACCATATGAGGCACGAAGCC 864

      240 .....ValAspSerSerCysSerGlyGlnProIleHisLeuTrpGl 253
                :::   |||   |||   |||            |||   ::
      865 TCCTTGGCAATCCGCTCCCCCTGCCCTGGG.....CTGGAGGA 902

      253 nPheLeuLysGluLeuLeuLysProHisSerTyrGlyArgPheIleA 278
          :   ::::::      :::|||:::|||         |||   :::::::
      903 GGACATGGAAGCCTACGTGCTGCGGCCAGCGCTCCCGGGCACCATCATGT 952

      279 rg...TrpLeuAspLys.............GluLysGlyIlePheLys 299
          ::   :::|||:::::               :::|||||:::|||
      953 AGTGCTACCTCACCCGTGACAAGCACGGCGTGGACAAGGGCTTGTTCCCC 1002

      281 Ile.....GluAspSerAlaGlnValAlaArgLe 299
          :::           :::|||||   |||      ||
     1003 CTCTACTACCTCTACCTGGAGACCTCTGACAGCCTGCAGCGCTTCCTCTT 1052

      290 aTrpGlyIleArgLysAsnArg...ProAlaMetAsnLyr 302
          |   |||   ::::::::::|||      :::   ||||||
     1053 GGCTGGGCGAAAGAGAAGAAGGAGCAAAACTTCTAATTAC 1092

seq_name: /cgn2_6/ptodata/1/pcnseq/pcnseqnp/NA2000_DAT:AAA94636

seq_documentation_block:
ID   AAA94636 standard; cDNA; 1733 BP.
XX
AC   AAA94636;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Human TULP2 cDNA.
XX
KW   Human; TULP; neurosensory detect; retina; retinal dystrophy; testes;
KW   chromosome 19q; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             94..1656
FT                   /*tag= a
FT                   /product= "Human TULP2"
XX
PN   US6114502-A.
XX
PD   05-SEP-2000.
XX
PF   27-FEB-1998;    98US-0032465.
XX
PR   22-AUG-1996;    96US-0701480.
PR   04-SEP-1996;    96US-0706292.
PR   10-APR-1996;    96US-0630592.
PR   17-SEP-1996;    96US-0714991.
PR   30-APR-1997;    97US-0850218.
PR   01-AUG-1997;    97US-0904699.
PR   17-SEP-1997;    97US-0932406.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   North M,  Nishina P,  Noben Trauth K,  Naggert J;
XX
DR   WPI; 2000-586163/55.
DR   P-PSDB; AAB29966.
XX
PT   Mammalian proteins expressed in retina and brain, useful for producing
PT   antibodies and for diagnosing neurosensory defects including cochlear
PT   degeneration, peripheral retinal degeneration and cone rod retinal
PT   dystrophy -
XX
PS   Disclosure; Columns 63 64; 61pp; English.
XX
CC   The present sequence is the coding sequence for human TULP2. This
CC   sequence is a member of the neurosensory detect associated gene family,
CC   and is expressed in the retina and testes. TULP2 is tightly linked to
CC   framework marker WI 9028 on chromosome 19q. The present sequence is
CC   useful as an immunogen to raise antibodies that specifically identify
CC   TULP expressing cells and in drug screening assays directed at
CC   neurosensory defects. The protein encoded by the present sequence can be
CC   used for the treatment of neurosensory degenerative conditions (retinal
CC   dystrophies) e.g. retinitis pigmentosa, combined cone rod dystrophy, age
CC   related macular dystrophy, Stargardt's macular dystrophy, Best's disease,
CC   pigment pattern dystrophies, central alveolar choroidal dystrophy,
CC   dominant drusen, hereditary haemorrhagic macular dystrophy, North
CC   Carolina macular dystrophy, pericentral choroidal dystrophy, adult
CC   foveomacular dystrophy, benign concentric annular macular dystrophy,
CC   central aureolar pigment epithelial dystrophy, congenital macular
CC   coloboma, dominantly inherited cystoid macular oedema, familial foveal
CC   retinoschisis, fenestrated sheen macular dystrophy, progressive foveal
CC   dystrophy, slowly progressive macular dystrophy, Sorsby's
CC   pseudoinflammatory dystrophy, progressive cone dystrophy, Leber's
CC   congenital amaurosis and Goldman-Favre syndrome.
XX
SQ   Sequence 1733 BP; 449 A; 480 C; 455 G; 349 T; 0 other;


alignment_scores:
            Quality:  108.50              Length:     414
              Ratio:  0.764                 Gaps:      17
 Percent Similarity:  45.223   Percent Identity: 22.611

alignment_block:
 US-09-126-945B-2 x AAA94636   ..

 Align seq 1/1  to: AAA94636  from: 1  to: 1733

       31 AlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPr 47
          :::|||:::|||||              :::|||   |||:::|||   ||
      275 TCTGCGGGAGGAGCGCCTTTAGGTGACAGAGCCTTGGGAACCTTTCT 322

       47 oProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheA 64
          ||||           ::::||   ||||||||:
      323 TCCGGAACAAAGTGTCAGAGGCACATCTGCCCTC.............. 356

       64 spMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGly... 79
                                    |||      ::::::||||||
      357 .....TGGCATCTACAGTGGCCTGGGCAC 380

       80 .................AlaSerSerArgGluGluProProGluGlnPr 90
                           :::::::|||   :::||||||||:::
      381 CGTGAGCTGTGGTGGAGACGGCAGGGGCGAGCGCGGCCTCCCGACACCGC 430

       90 oGluGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspL 107
          ::::::     |||   :::|||::::::|||:::     |||   |
      431 GGACAGAAGCAGTGTTCAGGAATCTCGGTCTCCAGTCCCCTTTGTTATCC 480

      107 eu.ValProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGl 123
          || :::|||   :::         |||         ||||||::|||
      481 TGGCTCGCAGACAATTCCGATGCAGAA.....TTGGAGGAAGTC.. 519

      123 nSerMetValValGlyGluValLeuLysAspIleGluThrAlaCysLysL 140
```

```
                                    ::::::|||:::::::
     520 ..............................TCCGTGGAGAATGGT....... 534

     140 euLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGln 156
             :::::::::       |||:::      ||||||          :::
     535 .....TCCGTCTCTCCCCCACCTTTTAAACAGTCTCCGAGAATCCGACGC 579

     157 LysTrpLeuLeuTrpThrGluHisGlnTyrArqLeuProProMetGlyLy 173
         |||           |||       ||||||             |||      ::
     580 AAGGGT......TGGCAAGCCCACCAA.........CGACCTGGCACCCG 614

     174 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGlqGlu 190
         ::||       |||      ::::::::::::            ::::::
     615 TGCAGAGGGTGAGAGTGACTCCCAGGATATGGGAGATGCACACAAGTCAC 664

     190 lnPheArqGlnArqSerPro...LeuGlyGlyAspValLeuHisAlaHis 205
                   :::|||   :::   ||||||   ::::::   :::
     665 CCAATATGGGACCAAACCCTGGAATGGATGGTGACTGTGTATATGAAAAC 714

     206 LeuAspIleTrpLysSerAlaAlaTrpMetLysGluArq........... 218
         |||   :::   ||:::          ||:::|||
     715 TTGGCCTTCCAAAAGGAAGAAGACTTGGAAAAGAAGAGAGAGGCCTCTGA 764

     219 ..........ThrSerProGlyAlaIleHis.........TyrCysAlaS 229
                   |||:::   :::|||   |||              :
     765 GTCTACAGGGACCAACTCCTCAGCAGCACACAACGAAGAGTTGTCCAAGG 814

     229 erThrSerGlnGluSerTrpThrAspSerGln................. 259
         ::   :::   |||:::   ||||||||||:::
     815 CCCTGAAAGGGCAGGGTGGCACGGACAGCGACCATATGAGGCACGAAGCC 864

     240 .........ValAspSerSerCysSerGlyGlnProIleHisLeuTrpGl 253
                 :::   |||   |||   |||             |||   ::
     865 TCCTTGGCAATCCGCTCCCCCTGCCCTGGG............CTGGAGGA 902

     254 nPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArqPheIleA 270
         :   ::::::      :::|||:::|||        |||   :::::::
     903 GGACATGGAAGCCTACGTGCTGCGGCCAGCGCTCCCGGGCACCATGATGC 952

     270 rg...TrpLeuAsnLys..............GluLysGlyIlePheL,.. 289
         ::   :::|||:::::              :::|||||:::|||
     953 AGTGCTACCTCACCCGTGACAAGCACGGCGTGGACAAGGGCTTGTTCCCC 1002

     281 Ile.....................GluAspSerAlaGluValAlaArqLe 298
         :::                  :::||||||   |||         ||
    1003 CTCTACTACCTCTACCTGGAGACCTCTGACAGCCTGCAGCGCTTCCTCCT 1052

     290 nTrpGlyIleArqLysAsnArq...ProAlaMetAsnTyr 302
         |   |||   ::::::::::||       :::   ||||||
    1053 GGCTGGGCGAAAGAGAAGAAGGAGCAAAACTTCTAATTAC 1092

seq_name: /SIBIOS/seqdata/geneseq/geneseqp/NA2000_DAT:AAF21305

seq_documentation_block:
ID   AAF21305 standard; DNA; 6542 BP.
XX
AC   AAF21305;
XX
DT   14 MAR 2001  (first entry)
XX
DE   Human low adenosine antisense oliqonucleotide related sequence #2872.
XX
KW   low adenosine antisense oliqonucleotide; phosphorothioate; allerqy;
KW   human; airway disorder; bronchoconstriction; lunq inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analqesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allerqic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736 A2.
XX
PD   26-OCT-2000.
XX
PF   24 MAR 2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC ) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679589/66.
XX
PT   Low adenosine (A) content antisense oliqonucleotides which do not
PT   triqqer adenosine receptors durinq metabolism, useful e.q. for treatinq
PT   cancers and respiratory obstructions
XX
PS   Disclosure; Page 1293-1295; 1592pp; Enqlish.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oliqonucleotides and compositions (I) comprising them. In the antisense
CC   oliqonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analqesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oliqonucleotides and (I) can be used to down-requlate the
CC   expression and or activity of target polypeptides associated with
CC   lunq/respiratory disorders and malignancies, such as stimulatinq and
CC   activatinq peptide factors and transmitters, transcription factors,
CC   immunoqlobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endoqenously produced specific and non specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, qrowth factors, vasoactive peptides and
CC   receptors, bindinq proteins and malignancy associated proteins. The
CC   antisense oliqonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lunq inflammation, allerqy(ies)
CC   and/or surfactant hyperproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allerqies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allerqic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fraqments and antisense oliqonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 6542 BP; 1511 A; 1824 C; 1790 G; 1407 T; 0 other;

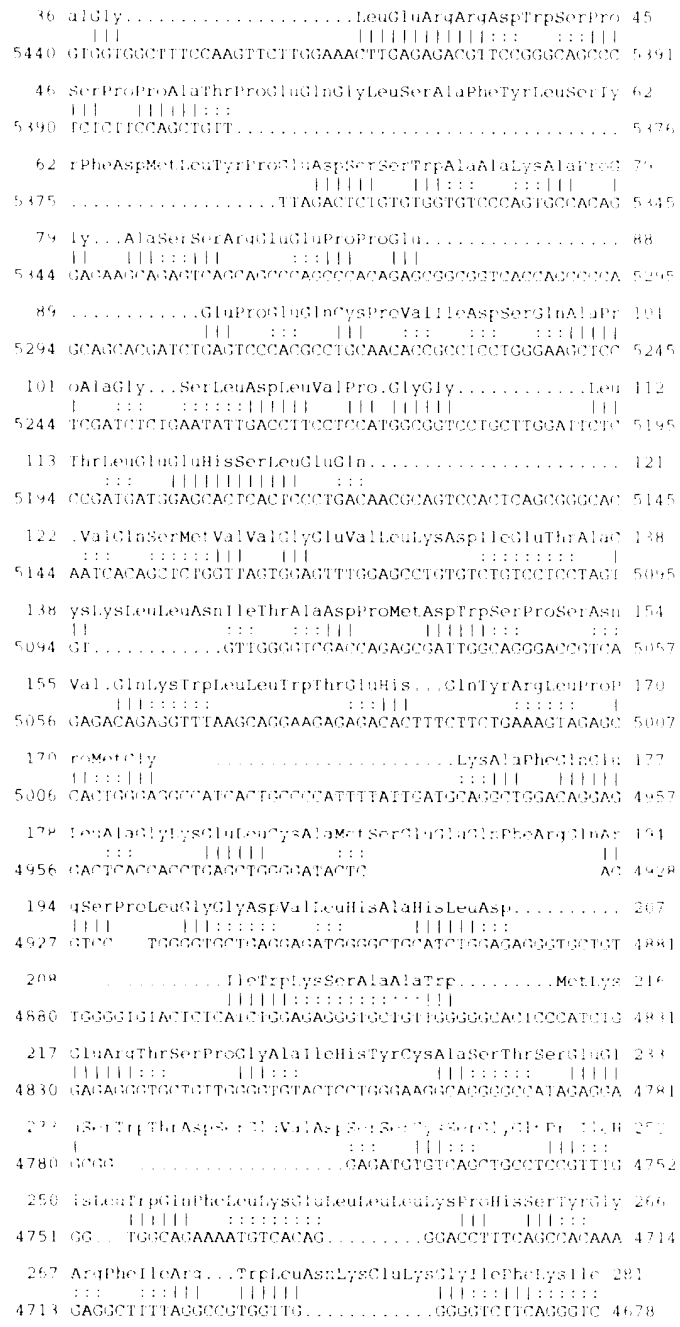

alignment_scores:
            Quality:  107.00              Length:     534
              Ratio:  0.709                 Gaps:      24
 Percent Similarity:  45.210   Percent Identity: 25.150

alignment_block:
 US-09-126-945B-2 x AAF21305/rev   ..

 Align Seq 1/1  to reverse of: AAF21305  from: 1  to: 6542

       5 SerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProProAs 21
         ||||||   :::   |||:::|||       :::|||   :::|||||||
    5534 AGTCCTGAGATT...TCAATCTCTGCTGTTCAGCTCGAAGTTCCTCCAGT 5488

      21 pThrValSerArq......ThrGlyLeuGluLysAlaAlaAlaGlyAlaV 46
            |||:::          |||||||:::   :::|||   :::|||
    5487 TCCTGTGAAATCTGCGTGGACTGGGATT...TCTGCGCCATCTGGTCACC 5441
```



seq_name: /SIEVE1:pdb1/Genseq1/Genseqp_NA2000.DAT:AAA35183

seq_documentation_block:
ID AAA35183 standard; DNA; 6532 BP.
XX
AC AAA35183;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:57.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers
XX
PS Disclosure; Page 1210-1212; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35412 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which corresponds to SEQ ID NO:1 to 2615, and thus the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 6532 BP; 1511 A; 1824 C; 1790 G; 1407 T; 0 other;

alignment_scores:

```
           Quality:  107.00              Length:     334
             Ratio:  0.709                 Gaps:      24
Percent Similarity:  45.210   Percent Identity: 25.150

Alignment block:
US 09 126 945B 2 x AAA35183/rev   ..

Align seq 1/1  to reverse of: AAA35183  from: 1  to: 6532

       5 SerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProProAs 21
         ||||||   :::   |||:::|||       :::|||   :::|||||||
    5534 AGTCCTGAGATT...TCAATCTCTGCTGTTCAGCTCGAAGTTCCTCCAGT 5488

      21 pThrValSerArg......ThrGlyLeuGluLysAlaAlaAlaGlyAlaV 36
             |||:::          |||||||:::   :::|||   :::|||
    5487 TCCTGTGAAATCTGCGTGGACTGGGATT...TCTGCGCCATCTGGTCACC 5441

      36 alGly.....................LeuGluArgArgAspTrpSerPro 45
           |||                     ||||||||||||||:::   :::|||
    5440 GTGGTGGCTTTCCAAGTTCTTGCAAACTTGAGAGACCTTCCGGCCAGCCC 5391

      46 SerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTy 62
         |||   ||||||:::
    5390 TCTCTTCCAGCTGTT................................... 5376

      62 rPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProG 79
                               ||||||   |||:::   :::|||   |
    5375 ...................TTAGACTCTGTGTGGTGTCCCAGTGCCACAG 5345

      79 ly...AlaSerSerArgGluGluProProGlu.......          88
         ||   |||:::|||       :::|||   |||
    5344 GAGAAGCAGAGTCAGCAGCCCAGCCCACAGAGCGGCGGTCACCACCCCCA 5295

      89 ............GluProGluGlnCysProValIleAspSerGluAlaPr 101
                    |||   :::   |||   :::   :::   :::||||||
    5294 GCAGCACGATCTGAGTCCCACGCCTGCAACACCGCCTCCTGGGAAGCTCC 5245

     101 oAlaGly...SerLeuAspLeuValPro.GlyGly.............Leu 112
         |   :::   :::::||||||   ||| ||||||               |||
    5244 TCGATCTCTGAATATTGACCTTCCTCCATGGCGGTCCTGCTTGGATTCTC 5195

     113 ThrLeuGluGluHisSerLeuGlnGln...................... 121
           :::   ||||||||||||   :::
    5194 CCGATGATGGAGCACTCACTCCCTGACAACGCAGTCCACTCAGCGGGCAC 5145

     122 .ValGlnSerMetValValGlyGluValLeuLysAspIleGluThrAlaC 138
          :::   ::::::|||   |||                 :::::::::   |
    5144 AATCACAGCTCTGGTTAGTGGAGTTTGGAGCCTGTGTCTGTCCTCCTAGT 5095

     138 ysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsn 154
         ||              :::   :::|||       |||||||:::      :::
    5094 GT............GTTGGGGTCGACCAGAGCGATGGCAGGGACCGTCA 5057

     155 Val.GlnLysTrpLeuLeuTrpThrGluHis...GluTyrArgLeuProP 170
           |||::::::                 ::|||                |||   |
    5056 GAGACAGAGGTTTAAGCAGGAAGAGACACACTTTCTTCTGAAAGTAGAGC 5007

     170 roMetGly.........................LysAlaPheGlnGlu 177
         ||:::|||                       :::|||   ||||||
    5006 CACTGGGAGGCCATCACTGCCCCATTTTATTGATGCAGGCTGGACAGGAG 4957

     178 LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnAr 194
         :::     ||||||      :::                        ||
    4956 GACTCACCACCTGAGCTGGGGATACTC.....................AG 4928

     194 qSerProLeuGlyGlyAspValLeuHisAlaHisLeuAsp.......... 207
         ||||     |||:::::   :::      |||||||:::
    4927 GTCC...TGGGGTGCTGAGGAGATGGGGCTGCATCTGGAGAGGGTGCTGT 4881

     208 ..............IleTrpLysSerAlaAlaTrp........MetLys 216
                     ||||||:::::::::::::|||
    4880 TGGGGTGTACTCTCATCTGGAGAGGGTGCTGTTGGGGGCACTCCCATCTG 4831

     217 GluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGl 233
         ||||||||:::      |||:::            |||::::::   |||||
    4830 GAGAGGGTGCTGTTGGGGTGTACTCCTGGGAAGGCAGGGGCCATAGAGGA 4781

     233 uSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleH 250
         |                          :::   |||:::      |||:::
    4780 GCGG.....................GAGATGTGTCAGCTGCCTCCGTTTG 4752

     250 isLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGly 266
           ||||||   :::::::::            |||   ||||::
    4751 GG...TGGCAGAAAATGTCACAG.........GGACTTTCAGCCACAAA 4714

     267 ArgPheIleAra...TrpLeuAsnLysGluLysGlyIlePheLysIle 281
         :::   :::|||   ||||||            |||:::|||::::::
    4713 GAGGCTTTTAGGCCGTGGTTG............GGGGTCTTCAGGGTC 4678

Seq_name: 'US-09-126-945B-2 x AAA35183/rev'; Score=pg:NA2000:SAT.AAC75875

Seq_documentation_block:
ID   AAC75875 standard; cDNA; 1029 BP.
XX
AC   AAC75875;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF1430 polynucleotide sequence SEQ ID NO:2859.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000;   2000US-0540764.
XX
PA   (CURA ) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB41666.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease
XX
PS   Claim 5; Page 2099 2100; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
```

```
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic, hypotensive, dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 1029 BP; 250 A; 325 C; 224 G; 229 T; 1 other;


alignment_scores:
            Quality:  105.00              Length:     299
              Ratio:   0.802               Gaps:      12
 Percent Similarity:  43.813   Percent Identity: 21.405

alignment_block:
 US-09-126-945B-2 x AAC75875   ..

 Align seq 1/1  to: AAC75875  from: 1  to: 1029

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
          ::::::||||||   ||||||                 |||:::   :::
       22 GTCTCCTCCGCGCCCCCGGGA..............CCCAACACAATGCCA 56

       17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 34
             ||||||:::::::::::::                 ||||||||||:
       57 CCAGCCGCCTGAGTCCACCGCC..............GCCGCCGCCG 88

       34 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaIh 50
          ::|||:::   ::::::   |||      :::      |||
       89 CCGCTGCAGACATTAGCGCTAGGAAGATGGCG...CACCCGGCAATGTTC 135

       51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeu.. 66
          |||:::::|||   ::::::          |||   ::::::
      136 CCTCGAAGGGGCAGCGGTAGTGGCAGCGCCTCTGCTCTCAATGCAGCAGG 185

       67 .....TyrProGluAspS 71
                          ::|||     |
      186 TACCGGCGTCGGTAGTAATGCCACATCTTCCGACGATTTTCCGCCTCCGT 235

       71 erSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluProPro 87
          ||           |||:::||||||||   :::   |||
      236 CGCTGCTTCAGCCGCCGCCCCCTGCAGCATCTTCTACGTCGGGACCACAG 285

       88 GlyGluProGluGlnCysProValIleAspSerGlnAlaProAlaGlySe 104
             |||   |||         :::   ||||||||         ::
      286 CCTCCGCCTCCACAAAGCCTGAACCTCCTTTCGCAGGCTCAGCTGCAGGC 335

      104 rLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGlu 121
          :     |||:::||||||||||     :::::::::   |||      |
      336 ACAGCCTCTTGCGCCAGGCGGAACTCAAATGAAAAAGAAAAGTGGCTTCC 385

      121 lnValGlnSerMetValValGlyGluValLeuLysAspIleGluThrAla 137
          |:::   |||::::::   :::::::::      :::|||:::::::
      386 AGATAACTAGCGTTACTCCTGCTCAGATCTCCGCTAGTATCAGCTCTAAC 435

      138 CysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAs 154
             ::::::   :::   |||   :::      |||   :::   |||::
      436 AACAGTATAGCAGAGGACACTGAGAGCTATGATGATCTGGATGAATCTCA 485

      154 nValGluLysTrpLeuLeuTrpThrGluHisGlnIyrArgLeuProProM 171
          ::::
      486 CACG..................................................... 489
```

```
      171 etGlyLysAlaPheSerGlnGlnLeuAlaGlyLysGluLeuLysAlaMetSer 187
                    ::::::|||::::::::::|||:::      :::|||
      490 ..............GAACATCTCTCTTCTCGGAGATCCTTGATGTGTCA 525

      188 ..........................GluGln..GluPheArgGlnArg 194
                                      |||  |||:::   :::|||
      526 CTTTCCAGGGCTACTGACTTAGGGGAGCCCGAACCCAGCTGCTCAGAAGA 575

      195 SerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleArgLysSe 211
          ::|||               |||   :::      |||   |||   ||
      576 GACCCT.........AAATAACTTCCAGGAAGCCGAGACACCTGGGGCAG 616

      211 rAlaAlaIrpMetLysGluArgIhrSerProGlyAlaIleHisIyr.... 226
          |      :::::::::   :::   |||:::::      :::
      617 T.........CTCTCCCAACCAGCCCACCTTCCTCAGCCTCATTTGCCT 657

      227 ..................................CysAlaSerIhr... 240
                                           |||::::|||||
      658 CACCTTCCACAACAGAATGTTGTGATCAATGGGAATGCTCATCCACACCA 707

      241 ..............SerGluGlnSerIrpIhrAspSerGluVal..... 260
                       |||:::::|||||:::      :::
      708 CCTCCATCACCACCATCAGATTCATCATGGGCACCACCTCCAACATGGTC 757

      241 ..........AspSerSerCysSerGlyGlnProIleHisLeuIrp 252
                     |||   |||   ||||||   ||||||   |||
      758 ACCACCATCCATCTCATGTTGCTGTGGCCAGTGCATCCATTACTGG 803
```

```
seq_name: /cgn2_6/ptodata/1/pcds/pcds_NAC1/1251_AAC91988

seq documentation block:
ID   AAC91988 standard; cDNA; 1608 BP.
XX
AC   AAC91988;
XX
DT   20-MAR-2001  (first entry)
XX
DE   Human T-bet coding sequence.
XX
KW   Human; T bet; T box expressed in T cells protein; autoimmune disease;
KW   Th2-type dysfunction; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073453 A1.
XX
PD   07-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US15345.
XX
PR   02-JUN-1999;   99US-0137085.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Glimcher LH,  Szabo SJ;
XX
DR   WPI; 2001-061537/07.
DR   P-PSDB; AAB51013.
XX
PT   Novel protein, T-bet, a new member of T box family of transcription
PT   factors is useful for identifying modulators of T-bet activity which
PT   treat allergies, cancers, autoimmune and infectious diseases
XX
PS   Claim 2; Page 116-118; 126pp; English.
XX
CC   The present sequence is the coding sequence for human T box expressed in
CC   T cells (T-bet) protein. T-bet can be used in the treatment of autoimmune
CC   diseases associated with a Th2-type dysfunction e.g. diabetes mellitus,
CC   arthritis, multiple sclerosis, autoimmune thyroiditis, psoriasis,
CC   keratoconjunctivitis, alopecia areata, allergic responses due to
CC   arthropod bite reactions, Crohn's disease, aphthous ulcer, iritis,
CC   conjunctivitis, keratoconjunctivitis, ulcerative colitis, asthma,
```

```
CC   allergic asthma, cutaneous lupus, erythematosus, scleroderma, vaginitis,
CC   proctitis, and drug eruptions. The role of T bet is to drive
CC   differentiation of Th cells.
XX
SQ   Sequence 1608 BP; 411 A; 541 C; 481 G; 285 T; 0 other;


alignment scores:
            Quality:  105.00              Length:     478
              Ratio:  0.625                 Gaps:      20
 Percent Similarity:  44.444   Percent Identity: 21.693

alignment block:
 DS 09 126 945B 2 x AAC91988  ..

 Align seq 1/1  to: AAC91988  from: 1  to: 1608

       1 MetGlySerAlaSerProGlyLeuSerSerVal..................... 11
         ||||||   ::::::||||||   :::::::::
       1 ATGGGCATCGTGGAGCCGGGTTGCGGAGACATGCTGACGGGCACCGACGCC 50

      12 ..........................................SerProSerHisL 16
                                                  :::|||:::|||
      51 GATGCCGGGGAGCGACGAGGGCCGGGCGCCTGGCGCCGACCCGCAGCACC 100

      16 euLeuLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAla 32
           :::   |||:::      ::::::   :::   |||:::   :::
     101 GCTACTTCTACCCGCAGCCGGGGCGCAGGACGGCGACGACGTCCGCG 150

      33 AlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPro..... 47
         :::|||:::::|||                ||||||   |||
     151 GGCGGCAGCCTGGGG..................TCTCCCTACCCGGGGGG 192

      48 ..........ProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuS 61
                   ||||||   |||::::::   ||||:::|||:::
     193 CGCCTTGGTGCCCGCCCCGCCGAGCCGCTTCCTTGGAGCCTAC........ 229

      61 erTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla 77
                        ||||||         :::   ||||||
     226 ..............GCCTACCCCCCGCACCCCAG...GCGGCCGGCTTC 258

      78 ProGlyAlaSerSerArgGluGluProProGluGluProGlu...GlnGly 93
         ||||||||||::::::          ||||||   :::   |||
     259 CCCGGCGCGGGCGAGTCCTTCCCGCCGCCCGCGGACGCCGAGGGCTACCA 308

      93 sProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProG 110
          |||                ||||||                |||   ||||
     309 GCCGGGCGAGGCTACGCCGCCCCGGACCCGCGCGCCGGGCTCTACCCGC 358

     110 lyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetVal 126
         ||       |||::::::::::|||
     359 GG......CCGCCTGAGGACTACGCGCTACCC.................. 384

     127 ValGlyGlnValLeuLysAspIleGluThrAlaCysLysLeuAsnLeuA 143
                          :::|||:::::::   ||||||
     385 ..............GCGGGACTGGAGGTGTCGGGCAAACTGAGGGTGGC 419

     144 eThrAlaAspProMetAspTrpSerProSerAsnVal...GlnLysTrpL 159
           :::   :::   ||||||      |||      |||      :
     420 GCTCAACAACCACCTGTTGTGTCCAAGTTTAATCAGCACCAGACAGAGA 469

     159 euLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPhe 175
         :::::   |||:::::      ||||::   |||:::      :::|||
     470 TGATCATCACCAAGCAGGGACGGCGGATGTTCCCATTC...CTGTCATTT 516

     176 GlnGluLeuAlaGlyLysGln..............................  182
           :::|||||||   |||
     517 ...ACTGTGGCCGGGCTGGAGCCCACCAGCCACTACAGGATGTTTGTGGA 563

     183 .LeuCysAlaMetSerGluGluGlnPheArgGlnArgSer..........  195
         :::       ::::::::::::::::::::|||   ::::||
     564 CGTGGTCTTGGTGGACCAGCACCACTGGGGGTACCAGAGCGGCAAGTGGG 613

     196 .......................ProLeuGlyGlyAspValLeuHisAla 204
                                  :::   ||:::   |||::::::
     614 TGCAGTGTGGCAAAGGCCGAGGGCAGCATGCCAGGAAACCGCCTGTACGTC 663

     205 HisLeuAspIleTrpLysSerAlaAla...TrpMetLysGluArgthrSe 220
         |||   |||       ::::::::::|||   ||||||::::::::::||
     664 CACCTGGACTTCCCAACACAGGAGCGCACTGGATGGCCAGGAAGTTTC 713

     220 rProGly.................................................. 222
         |   |||
     714 ATTGGGAAACTAAAGCTCACAACAACAAGGGGGCGTCCAACAATGTGA 763

     223 .....................AlaIleHis...TyrCysAlaSerThrSer 231
                              ::::::|||   |||
     764 CCCAGATGATTGTGCTCCAGTCCCTCCATAAGTACCAGCCCCGGCTGCAT 813

     232 GluGluSerTrpThrAspSerGluValAspSerSerCysSerGlyGluPr 248
           :::   :::|||:::|||   :::::::::|||::::::::::
     814 ATCGTTGAGTTGAACGACGGAGAGCCAGAGCCAGCCTGCAACGCTTCCAA 863

     248 oIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerT 265
           |||::::::   |||   :::   ::::::      :::|
     864 CACGCATATCTTTACTTCCAAGAAACCCAGTTCATGCCGTGACTGCCT 913

     265 yrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIle 281
         ||                      |||   |||          ::::||||
     914 AC..................CAGAATGCCGACATTACTCAGCTGAAAATT 945

     282 GluAspSerAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgPr 298
           ||::::          ||::::       |||:::|||:::|||
     946 ...GATAAATAACCCCTTTGCCAAA......GGATTCCGGGACAACTTTGA 986

     298 oAlaMetAsnTyrAspLysLeuSerArgSerIle 309
         :::|||   |||   :::::::::   ||||||
     987 GTCCAG...TACACATCTGTGGACACCAGCATC 1017

Seq name, "file%s": file:/home/.../pCNA1994.DAT:AAV12201

seq documentation block:
ID   AAV12201 standard; cDNA; 2448 BP.
XX
AC   AAV12201;
XX
DT   14-MAY-1998  (first entry)
XX
DE   Human neuronal nicotinic acetylcholine receptor beta 2 subunit cDNA.
XX
KW   Human; neuronal nicotinic acetylcholine receptor; beta 2 subunit;
KW   brain tissue; screening; NAChR; antibody; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/qualifiers
FT   CDS             265..1773
FT                   /*tag   a
FT                   /product  "neuronal nicotinic acetylcholine receptor
FT                             beta 2 subunit"
XX
PN   WO9420617-A2.
XX
PD   15-SEP-1994.
XX
PF   08-MAR-1994;   94WO-US02447.
XX
PR   08-MAR-1993;   93US-0028041.
XX
PA   (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PA   (SIBI ) SIBIA NEUROSCIENCES INC.
XX
PI   Elliott KJ,  Ellis SB,  Harpold MM;
XX
DR   WPI; 1994-304024/37.
DR   P-PSDB; AAW44157.
```

```
XX
PT   Human neuronal nicotinic acetylcholine receptor subunits and cDNA
PT   also transformed cells useful for screening cpds. which modulate
PT   activity of the receptor
XX
PS   Example 2; Page 82-83; 99pp; English.
XX
CC   The present sequence encodes a human neuronal nicotinic acetylcholine
CC   receptor (NAChR) subunit. The cells expressing the alpha and/or beta
CC   NAChR subunits may be used in a method of screening compounds to
CC   identify any which modulate the activity of human neuronal NAChR.
CC   Subunit specific antibodies may be used to monitor the distribution
CC   and expression density of various subunits in normal vs diseased brain
CC   tissues. Testing of single receptor subunits or specific receptor
CC   subunit combinations with a variety of potential agonists or antagonists
CC   provides information with respect to the function and activity of the
CC   individual subunits and should lead to the identification and design of
CC   compounds that are capable of very specific interaction with one or
CC   more receptor subtypes. The resulting drugs should exhibit fewer
CC   unwanted side effects than drugs identified e.g. screening with cells
CC   that express a variety of subtypes.
XX
SQ   Sequence 2448 BP; 457 A; 801 C; 705 G; 485 T; 0 other;


alignment_scores:
            Quality:  104.50              Length:     474
              Ratio:  0.661                 Gaps:      20
 Percent Similarity:  42.246   Percent Identity:  22.460

alignment_block:
 US-09-126-945B-2 x AAV12281/rev    ..

 Align seq 1/1   1: reverse of AAV12281  from 1  to 2119

        6 ProGlyLeuSerSerValSerProSerHisLeu..................  16
          |||||||  ||||          ||||||:::|||
     2242 CCAGGCCTCTCTCTTCCCCTCCCTTCAGAGCTACCATTGTGTCAGGCCAG 2193

       17 ................LeuLeuProProAspThrValSerArgThrGlyL  28
                          |||||||||          :::|||   |||   |
     2192 CCCGGTGCGGTGTGAGCTCCTGCCCTTCCACCCCTTGTCCTATACCCCTC 2143

       28 euGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSer  44
          ||           :::|||:::               |||:::
     2142 TTCTCCCATGACTTGGAGGCAGC..............AGACAATCCTGT 2108

       45 ProSerPro.........ProAlaThrProGluGlnGly...........  54
          |||   |||         ||||:::||||||:::   |||
     2107 CCCCTTCGTGGCCCCTTGCCCGTCACTCCTCAGGGTGCATGGCTCTGCCA 2058

       55 .............LeuSerAlaPhe.......................I  59
                      |||||:::::                        :
     2057 TCTCTGTTTCCACTTAGCAGCCTTGCTCCCCTCCTCAACACTATTGCCT 2008

       59 yrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAla  75
          ::   :::      |||          |||:::::::|||||||||||:::
     2007 GGGCTGGTACAAGAGATGGAGAGCAGCCAAAACAAAGCAGTTGGTCCAGC 1958

       76 LysAlaProGlyAlaSerSer.........ArgGluGlyProProGluGl  89
                 ||||||||:::             ||||::
     1957 CTCCAGGTTGGAGCTGCTGTCTGTGGATGGAGGGGTGGGGCAGGAGGGAC 1908

       89 uProGluGlnCysPro......ValIleAspSerGlnAlaProAlaGlyS 104
            ||:::::::::          :::      :::::::||||
     1907 TCCACATGAAGCAAAAGGATGGATCTGTGGGGGCAGCGCCCT....... 1865

      104 erLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGlu 120
                 :::||||||||:      ::::::
     1864 ..........CTTCCTGGTAGCTCACTCGTCCATCCTCCACCCAACACT 1820

      121 GlnValGlnSerMetValValGly...........GluValLeuLysAs 133
             ||||||:::   |||:::|||           ||||:   |||
```

```
     1825 ACTGTGCAGCAGAGGCTGGCAGGCTGAAAGAGCATGCAGATGAGGAAGGG 1776

      133 pIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspI 150
            :::      :::      |||   :::::::::::           |
     1775 CCTCACTTGGAGCTGGGGGCTGAGTGGTCTGAGTGGAGGAAGGTGGTGGT 1726

      150 rpSerProSerAsnValGln.............lysTrpLeuLeuTrp 161
          ||      ||:::   :::                 ::::||
     1725 GGTGTAGTTCTGGAAGATAAGCTCCAGGAACATGCCGATGGTGC..... 1682

      162 ThrGluHisGluTyrArgLeuProProMetGlyLysAlaPheGlnGlnIle 178
            :::||:::   |||          |||:::::   ::::::
     1681 CAAAGACACAGACAAAGACAAAGATCCAGAGGAAGAGGCCGCCATCAGC 1632

      179 nAlaGlyLysGlnLeuCysAlaMetSerGluGluGlnPheArgGluArgS 195
                :::                |||:::::::     |||   ||||
     1631 ATCGCCACCTACTTCCAGTCCTCACTCACACCTCTGGTCATCTCCTCCGCT 1582

      195 erProLeuGly.GlyAspValLeuHisAlaHisLeuAspIleTrpLysSe 211
          ||      |||    |||      |:: ||
     1581 CCGCATGTGGTCTGCGATGAAGCGCACGCC..................... 1552

      211 rAlaAlaTrpMetLysGluArgThr.................SerProG 222
           :::     :::  |||  |||                   :::||||
     1551 .GTCCACCGCCTCCCGGAGGCCACAGCCACACGGCTCCCCTGAGCGCCCG 1503

      222 lyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSer 238
          |||||   |||:::|||      :::   |||
     1502 GGGCCCGCCACTGGTGCAGGCTCAGCCCCGAAGGC.............. 1468

      239 GluValAspSerSerCysSerGlyGlnProIleHisLeu............ 251
                                 ||||||||||:::|||
     1467 ................CCCGGCCAACCCTGCACCGACCGCGGGTTGA 1436

      252 ........................TrpGlnPheLeuLysGln....... 257
                                  |||   |||   :::|||
     1435 CGAAGCACGTGCAGGAGTCGGCCCCTGGGGCTTCGCGGAAGAAGAGGGCT 1386

      258 .................LeuLeuLeuLysProHisSerTyr......... 265
                          |||   :::|||:::::::
     1385 CCAGCGCCCTCGCGCTCACGCTGGCGTCGCCGCAGGCGCAGGCGCTGACG 1336

      266 GlyArgPheIleArgTrpLeu......AsnLysGluLysGlyIlePheLy 280
          |||   ::::::  ||||||          :::::::::::::|||   :::
     1335 GGCGCAATGATGGCGTGGCTGCTGCATGAAGAGCAGCGCGGGCAGCTTCT 1286

      280 sIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnA 297
            ||||||||:::   :::   ::::||      |||::::|||   :::|
     1285 CCAGGAAGACGACGTTCACCCAGGCGCC   CATGGTGTGCGTGGTGGGC 1239

      297 rgProAlaMetAsnTyrAsp 308
          ||   :::::::::   :::
     1238 GAGCGGTGGTGCACGTTGAG 1219

seq_name  _351-09_1_pda_1:Gross_proc:seq_NA1997_DAT:AAI48240

seq_documentation_block:
ID   AAI48240 standard; DNA; 2448 BP.
XX
AC   AAI48240;
XX
DT   09-APR-1997  (first entry)
XX
DE   Neuronal nicotinic acetylcholine receptor beta-2 subunit DNA.
XX
KW   Neuronal nicotinic acetylcholine receptor; nAChR; neurotransmitter;
KW   ligand-gated receptor; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             265..1773
FT                   /*tag   a
```

```
XX
PN   WO9641876 A1.
XX
PD   27 DEC 1996.
XX
PF   07 JUN 1996;   96WO US09775.
XX
PR   07 JUN 1995;   95US 0484722.
XX
PA   (SIBI ) SIBIA NEUROSCIENCES INC.
XX
PI   Elliott KJ, Harpold MM;
XX
DR   WPI; 1997-065464/06.
DR   P-PSDB; AAW09026.
XX
PT   Nucleic acids encoding nicotinic acetylcholine receptor sub units
PT   used in screening to determine the effect of drugs on the receptor
XX
PS   Disclosure; Page 75-77; 108pp; English.
XX
CC   A DNA sequence (AAT48240) codes for the beta 2 subunit (AAW09026) of
CC   the human neuronal nicotinic acetylcholine receptor (nAChR). Host
CC   cells, esp. mammalian cells or amphibian oocytes, carrying beta 2
CC   nucleic acids, opt. in combination with other alpha and/or beta
CC   subunit nucleic acids (see also AAT48242-49, AAT48241), express
CC   recombinant nAChR subunits useful for identifying cpds. that
CC   modulate the activity of human nAChRs.
XX
SQ   Sequence 2448 BP; 457 A; 801 C; 705 G; 485 T; 0 other;


Alignment scores:
            Quality:  104.50              Length:     474
              Ratio:  0.661                 Gaps:      20
 Percent Similarity:  42.246   Percent Identity: 22.460

Alignment block:
 US-09-126-945B-2 x AAT48240/rev   ..

 Align seq 1/1  to reverse of: AAT48240  from: 1  to: 2448

        6 ProGlyLeuSerSerValSerProSerHisLeu..................  16
          ||||||||||||          ||||||:::|||
     2242 CCAGGCCTCTCTCTTCCCCTCCCTTCAGAGCTACCATTGTGTCAGGCCAG 2193

       17 ..................LeuLeuProProAspThrValSerArgThrGlyL  28
                           |||||||||          :::|||   |||   |
     2192 CCCGGTGCGGTGTGACGCTCCTGCCCTTCCACCCCTTGTCCTATACCCCTC 2143

       28 euGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSer  44
          ||           :::|||:::          ||::::
     2142 TTCTCCCATGACTTGGAGGCAGC..............AGACAATCCTGT 2108

       45 ProSerPro.........ProAlaThrProGluGlnGly...........  54
          |||   |||         |||:::|||||:::   |||
     2107 CCCCTTCCTGGCCCCTTGCCCGTCACTCCTCAGGGTGGATGGCTCTGCCA 2058

       55 ............LeuSerAlaPhe.......................T  59
                     ||||||::::::                       :
     2057 TCTCTGCTTCCACTTAGCAGCCTTGCTCCCCTCCTCAACACTATTGCCT 2008

       59 yrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAla  75
          ::   :::       |||         |||:::::::||||||||:::::
     2007 GGGCTGGTACAAGAGATGGAGAGCAGCCAAAACAAAGCACTTGGTCCAGC 1958

       76 LysAlaProGlyAlaSerSer.........ArgGluGluProProGluGl  89
                 ||||||:::            ||:::
     1957 CTCCAGCTTGGAGCTGCTGTGTGTGGATGGAGGCGTGGGGGAGGAGGGAC 1908

       89 uProGluGluCysPro.....ValIleAspSerGlnAlaProAlaGlyS 104
          |||:::::::::        :::      ::::::||||||
     1907 TCCAGATGAAGCAAAAGGATGGATCTGTGGGGGCAGCGCCCCT....... 1865

      104 erLeuAspLeuValProGlyCysLeuThrLeuGluHisSerLeuGlu 120
                ::::||||||:::      ::::::
     1864 ..........CTTCCTGGTAGCTCACTCGCCATCCTCCACCCAACACT 1826

      121 GlnValGlnSerMetValValGly.............GluValLeuLysAs 133
          |||||||||:   ||||:::|||              ||||::   |||
     1825 ACTCTGCAGCAGAGGTGGCAGGGTGAAAGAGCATGGAGATGAGGAAGGG 1776

      134 pIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspI 150
          :::       :::       |||   ::::::::::::          |
     1775 CCTCACTTGGAGCTGGGGGCTCAGTGGTCTGACTGGAGGAAGCTGGTGGT 1726

      150 rpSerProSerAsnValGln..............LysTrpLeuLeuTrp 161
          ||      ||::::   :::              :::|||
     1725 GGTGTAGTACTGGAAGAGAGCTGCAGGAACATGCCGATGGTGC...... 1682

      162 IhrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGluGluLe 178
          :::||||::   |||              ||||::::::   ::::::
     1681 CAAAGACACACACAAAGACAAAGATCCAGAGGAAGAGGCGGTGCGATCACC 1632

      178 uAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGluArgS 195
                :::                  ||||::::::      |||   ||||
     1631 ATGGCGACGTACTTCACATCTCACTCACGCTCTGGTCATCGTCCTCGCT 1582

      196 erProLeuGlyGlyAspValGluHisAlaHisLeuAspIleTrpLysS 211
          ||      |||    ||      ||||||
     1581 CCGCATGTGGTCTGCGATGAAGCGCACGCC..................... 1552

      211 rAlaAlaTrpMetLysGluArgThr...................SerProG 222
            :::       :::   |||   |||                 ::||||
     1551 .GTCCACCGCCTCCTGGAGGCCACATGCCATACGGCTCCCTGAGCGCCG 1503

      222 lyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSer 238
          |||||   |||::|||       :::   |||
     1502 GGGCCCGCCACTGGTGCAGGCTCAGCCCCGAAGGC................ 1468

      239 GlyValAspSerSerCysSerGlyGlnProIleHisLeu............ 251
                               ||||||||||:::|||
     1467 ..................CCCGGCCAACCCCTGCACCGACCGCGGTTGA 1436

      252 ........................TrpGluPheLeuLysGlu........ 257
                                  |||   |||   :::|||
     1435 CGAAGCACGTGAGGAGTCGCCCCTGGGGCTTCGCGGAAGAAGAGGCT 1386

      258 ................LeuLeuLeuLysProHisSerTyr......... 265
                          |||   :::||||:::::::
     1385 CCAGCGCCCTCGCGCTCACGCTGGCGTCGCCGCAGGCGCAGGCGTGACG 1336

      266 GlyArgPheIleArgIrpLeu.......AsnLysGluLysGlyIlePheLy 280
          |||   ::::::   ||||||       :::::::::::::|||   :::
     1335 GGCGCAATGATGGCGTGGCTGCTGCATGAAGAGCAGCGCGGGCAGCTTCT 1286

      280 sIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnA 297
          ||||||:::   :::   :::|||       ||||:::|||   :::|
     1285 CCAGGAAGACGACCTTCACCCAGGGCGC...CATGGTGTGCGTGGTGGGC 1239

      297 rgProAlaMetAsnTyrAsp 308
          ||   :::::::::   :::
     1238 GAGCGGTGTGCACGTTGAG 1219

Seq name: gi|1781792|gb|AAA09042.1|(L27071) neuronal nicotinic acetylcholine receptor beta-2 subunit [Homo sapiens]

Seq documentation block:
ID   AAA09042 standard; cDNA; 2894 BP.
XX
AC   AAA09042;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Full length human TANGO 195 cDNA (clone T195AthHal70110).
XX
KW   TANGO 195; transmembrane protein; modulator; proliferation;
```

```
KW    differentiation; cell survival; CD2; Ig superfamily;
KW    gene therapy; anti-inflammatory; anti-allergic; cytostatic; ss.
XX
OS    Homo sapiens.
XX
FH    Key             Location/Qualifiers
FT    CDS             25..882
FT                    /*tag = a
FT    sig_peptide     25..87
FT                    /*tag = b
FT    mat_peptide     88..879
FT                    /*tag = c
XX
PN    WO200018800-A1.
XX
PD    06-APR-2000.
XX
PF    30-SEP-1999;    99WO-US22818.
XX
PR    30-SEP-1998;    98US-0163523.
XX
PA    (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI    Bousfield SJ;
XX
DR    WPI; 2000-293110/25.
DR    P-PSDB; AAY92163.
XX
PT    New cDNA molecules encoding human TANGO 191 and TANGO 195 proteins,
PT    useful for modulating cellular proliferation, differentiation or cell
PT    survival
XX
PS    Disclosure; Fig 9; 122pp; English.
XX
CC    Human TANGO 191 and 195 are transmembrane proteins, useful as modulating
CC    agents in a regulating a variety of cellular processes, e.g. cellular
CC    proliferation, differentiation and survival. Human TANGO 195 is a type 1
CC    transmembrane protein of the CD2 subgroup of the Ig superfamily. It may
CC    play a role in B cell leukaemia, immune response and autoimmune
CC    disorders, e.g. arthritis. The DNA sequence maps to human chromosome
CC    1q21. The TANGO 191 and TANGO 195 nucleic acids
CC    and proteins can be used in screening assays, detection assays and
CC    pharmacogenomics. The nucleic acids can be used in gene therapy and for
CC    detecting mRNA or genetic lesions. In particular TANGO 191 modulators
CC    are useful in treatment of immune and inflammatory disorders, e.g.
CC    asthma, graft-versus-host disease, rheumatoid arthritis, cancer and
CC    autoimmune diseases, e.g. myasthenia gravis.
XX
SQ    Sequence 2894 BP; 733 A; 818 C; 682 G; 661 T; 0 other;


alignment_scores:
            Quality:  104.50              Length:    339
              Ratio:  0.757               Gaps:       19
 Percent Similarity:  40.708   Percent Identity: 23.304

alignment_block:
 US-09-126-945B-2 x AAA09042   ..

 Align seq 1/1  to: AAA09042  from: 1  to: 2894

        2 GlySerAlaSerProGlyLeuSerSerVal..................... 11
          ||||||||||||||||   |||   |||
      132 GGCAGCCGTCCCCCTGGCTTCCAAGTCCGTGAGGCTATCTGGCGATCTC 181

       12 ................SerProSerHisLeuLeu......LeuProProA 21
                          ::::::::::::::::      ||||||   |
      182 TCTGGCCTTCAGAAGAGCTCCTGGCCACGTTTTTCCGAGGCTCCCTGGAG 231

       21 spThrVal.......................SerArgThrGlyLeuGlu 29
          ||:::|||                     |||
      232 ACTCTGTACCATTCCCGCTTCCTGGCCGAGCCCAGCTACACAGCAACCT 281

       89 LysAlaAlaAlaGlyAlaValGlyLeuGlnArgArgAsp.......... 12
          :::   :::|||   |||:::|||:::   ||||:::::
      282 CAGCCTGGAGCTGGGCCGCTGGAGTCTGGAGACAGCAGCAACCTCTCCG 331

       43 ..............................TrpSerProSerProProAlaI 50
                                       :::||||:::|||||||
      332 TGTTGATGGTGGACACAAGGGGCCAGCCCTGGACCCAGACCCTCCAGCT. 380

       50 hrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeu 66
                ||||||||:   :::
      381 .......CAAGGTGTACGATGC........................... 395

       67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 83
                                :::|||:::|   |||:::||   ::
      396 ..................AGTGCCAGGCCTGTGGTACAAGTGTTCA 424

       83 a.................GlnGluProProGlnGluProGlnGlnCysP 94
          :                :::          :::   |||
      425 TTGCTGTAGAAAGGGATGCTCAGCCTCCAAGACCTGCCAGGTTTTCTTG 474

       94 roValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGly 110
          ::::::      |||          :::::|||   ||||::   |||
      475 TCCTGTTGGGCCCCCAACATCAGCGAAATAACCTATAGCTGGCGACGGGA 524

      111 .........GlyLeu.ThrLeuGluGluHisSerLeu.........GluG 121
                  |||||    :::|||   ||||||||                 |
      525 GACAACCATGGACTTTGGTATGGAACCACACAGCCTCTTCACAGACGGAC 574

      121 luValGlnSerMetValValGlyGluValLeuLysAspIleGluThrAla 137
          |||||   |||:::   ::|||          :::|||:::      :::
      575 AGGTGCTGAGCATTTCCCTGGGACCAGGAGACAGAGATGTGGCCTATTCC 624

      138 CysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAs 154
          |||            |||:::::::::|||:::::|||:::   :::::
      625 TGC............ATTGTCTCCAACCCTGTCAGCTGGGACTTGGCCAC 662

      154 nValGlnLysTrpLeu......LeuTrpThrGluHisGlnTyrArg... 167
          :||||      ||| :::     :::      :::||||||   |||
      663 AGTCACGCCCTGGGATAGCTGTCATCATGAGGCAGCACCAGGGAAGGCCT 712

      167 ................................................. 167

      713 CCTACAAAGATGTGCTGCTGGTGGTGGTGCCTGTCTCGCTGCTCCTGATG 762

      168 ................LeuProProMetGlyLysAlaPheGlnGluLeuA 179
                          ||||||     |||       |||
      763 CTGGTTACTCTCTTCTCTGCCTGGCACTGGTGCCCCTGCTCAGGGA...A 809

      179 laGlyLysGlnLeuCysAlaMetSerGlnGlnGlnPheArgGlnArgSer 195
             |||::::::      :::::::|||          ||||||||:::
      810 AAAGAAAAAGGATGTCCATGCTGACAGAGTGGGTCCAGAGACAGAGAACC 859

      196 ProLeu.....................................GlyGl 199
          ||||||                                     |:|
      860 CCCTTGTGCAGGATCTGCCATAAAGGACAATATGAACTGATGCCTGGACT 909

      199 yAspVal......LeuHisAlaHisLeuAspIleTrpLys...SerAlaA 213
          :::|||      ||||||   |||     :::|||      :::::
      910 ATCAGTAACCCCACTGCACAGCACACGATGCTCTGGGACATAACTGGTG 959

      213 laTrpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSer 229
          |||||   :::     :::|||          :::   |||   :::
      960 CCTGGAAATCACCATGGTGCTCATATCTCCCATGGGAATCCTGTCCTGCC 1009

      230 ThrSerGluGluSerTrpThrAspSerGluValAspSerSerCysSerGl 246
          ::::::|||:::                                    ||
     1010 ICGAAGGAGCAG.....................................CCTGG 1026

      246 yGlnProIleHis 250
          |||||||   |||
     1027 GCAGCCATCACAC 1039

Seq1 name: /?1:38_5:51:Ra:*3b ... g_Na2000.DAT:AAC77550
```

```
Seq_documentation_block:
ID   AAC77560 standard; cDNA; 4105 BP.
XX
AC   AAC77560;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF3115 polynucleotide sequence SEQ ID NO:6229.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540764.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Loach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB43351.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease
XX
PS   Claim 5; Page 5413-5415; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
```

```
Sq   Sequence 4105 BP; 717 A; 918 C; 897 G; 571 T; 2 other;


alignment scores:
            Quality:  103.50              Length:     362
              Ratio:   0.784               Gaps:      14
  Percent Similarity:  36.464   Percent Identity: 19.337

alignment block:
 03-09-126-945B-2 x AAC77560/rc  ..

 Align seq 1/1  to reverse of: AAC77560  from: 1  to: 4105

        6 ProGlyLeuSerSerValSerProSerHisLeuLeuLeuProProAspTh 22
          ||||||:::::|||||      |||::::::                ||
     1670 CCAGGCATCGAGTCTGTCATTTCCTCAGAGGTTGGCGAGGGGTCTGGGAC 1621

       22 rValCysArgTyrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuG 39
          |||:::      :::::::::::::::      :::|||||||
     1620 GGTGACCTGAATGATGTCCAGATCCTCTGGCTCAATTTTGATCGGCCTC. 1572

       39 lnArgArgAspTrpSerProSerProProAlaThrProGluGlnGlyLeu 55
                        ||||||||||:::::||||||
     1571 .....AGCCCGCCCAGCTCCCCGGAG..... 1551

       56 SerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSe 72

     1551 ..................................................... 1551

       72 rTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGlnProProGluG 89
                     :::||||||                  |||:::::
     1550 ...........GTCCTGGC.................CCACAGT 1535

       89 lnProGlnThrCysProValIleAspSerGlnAlaProAla..... 102
          ::  ::::::  |||:::  |||          ||||||
     1534 CTTCAGACAGCTGCCCTTGTCTGACCGAGCATTCCCAGCAAGGTCAAGG 1485

      103 ...............GlySerLeuAspLeuValProGlyGlyLeuThrLe 114
                         |||  ::::::      ||||||||||:::
     1484 ACGGTGGCCCTAGAGACCCTCTGGAGAGGTGTCCTCTGGCGGGCTGGCTT.. 1437

      114 uGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluValL 131
                               :::|||:::||||||
     1436 .....GGCTCCAGCTTCGTGGTG.......... 1419

      131 euLysAspIleGlnThrAlaCysLysLeuLeuAsnIleThrAlaAspPro 147
                              :::|||:::|||:::        |||
     1418 .....TCTTTGGTAAACTTGTTC.....CCT 1398

      148 MetAspTrpSerProSerAsnVal.....GlnLysTrpLeuLeuTr 161
          :::          ||||||:::      :::||||||:::|||||
     1397 GTGAAAATTCGCTCATCAAACATCCTCTTAATGATGAAGTGGGATACTATG 1348

      161 pThrGluHisGlnTyrArg..... 167
          |   :::   :::   |||
     1347 GCGCTCCTCCAGGATCCGTTCAGCTTGGGGACTCCGTAAGTGCAGGGC 1298

      167 ..................................................... 167

     1297 GCTTGAAGGGGATCTTGTCCGGATGCCCACGATCTCCACAGCCTCCGGG 1248

      168 .....LeuProProMetGlyLy 173
                        :::   |||   ::
     1247 TCACAGGCAATCTTCCGGTAGGCACGGGACCATGTTGCCCAGGCCCAG 1198

      173 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGlnG 190
          :|||        |||          ::::::::|||:::   |||
     1197 GGCTTCGCATATCTGCTGTAAACAGGATCTGACACACTCC..... 1156

      190 lnPheArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu 206
                ||||||   |||   :::::|||
     1154 .....CGGAGCTGTCTCATGTCCTCGCTTTCTTTATGAAGGCG 1116
```

```
     297 AspIleTrpLysSerAlaAlaTrpMetLysGln..................  317
         :::   :::::::::   :::|||      |||
    1115 TCCCACTTCTCTGACTTGTCATGGACGATGGAGAAGAGAATGCGCTTGGA 1166

     318 ....................ArgThrSerProGlyAlaIleHisTyrCysA  339
                             |||      ||||||   :::   ::::|||
    1167 GGCATCGACGTTCTGGATGAGAGGCCCACCAGGCCCAGTGGGCTTCTGCT 1216

     339 IaSerThrSerGluGluSerTrpThrAspSerGluValAspSerSerCys  364
           :::   |||||||::|||||||:::         :::   ||||::
    1217 CACAACAGTCGGAGAAGTCTTGGGCACCGGTGGCCTTGGGCTCTGGCATG 1266

     365 SerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLy  391
         :::                                          ||||:
    1267 GGC.....................................AGACTCAG 1295

     391 sProHisSerTyrGly.................................A  407
         :|||::::||   |||
    1296 GCCCAGGTGGCTGGGGGCAAGGGGGCAGGAAGGTGCTTCCTGCTTGAGCT 1345

     407 rqPheIleArgArgLeuAsnLysGlu,..GlyIleIle,yeIleGluAsp  427
           :::   ||||||||||:::          |||:::
    1346 CTCGGATGGCGTGGTTGGGGAGCCCCGTGCTGTACAGGAAGCTGGGCATG 1395

     428 SerAlaGluValAla.....ArgLeuTrpGly 442
         :::   :::|||||          ||   ||:||
    1396 GAGGACTAAGTCGGCGGTTGGGGGCAGGTCCTGGGGT 1432

seq_name: US-09-126-945B-2.rng_seq_nos_seq_NA.res 661 AA_1* .5

seq_documentation_block:
ID   AAQ48024 standard; DNA; 997 BP.
XX
AC   AAQ48024;
XX
DT   21-APR-1994  (first entry)
XX
DE   AIGF gene #1.
XX
KW   cellular growth factor; androgen-induced growth factor; AIGF;
KW   cell growth; reagent; detection; anticancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             174..821
FT                   /*tag= a
FT                   /product= AIGF
XX
PN   JP05222096-A.
XX
PD   31-AUG-1993.
XX
PF   28-JAN-1992;   92JP-0038717.
XX
PR   28-JAN-1992;   92JP-0038717.
XX
PA   (SHIO ) SHIONOGI & CO LTD.
XX
DR   WPI; 1993-309194/39.
DR   P-PSDB; AAR41690.
XX
PT   Cell growth factor useful for detection of anticancer substance
PT   has specified amino acid sequence
XX
PS   Claim 1; Page 6-7 ; 10pp; Japanese.
XX
CC   The sequences given in AAQ48024-25 encode cellular growth factors, esp.
CC   androgen induced growth factor (AIGF).  The pref. protein of the
CC   invention comprises amino acids His35-Arg215 of sequence #1 or Gln23-
CC   Arg268 of sequence #2.  The AIGFs encoded by these sequences are
CC   useful as cell growth reagents and for the detection of anti-cancer
CC   substances
XX
SQ   Sequence 997 BP; 192 A; 330 C; 293 G; 182 T; 0 other;


alignment_scores:
            Quality:  103.000              Length:    178
              Ratio:  1.288                Gaps:        9
 Percent Similarity:  44.944   Percent Identity: 29.213

alignment_block:
 US-09-126-945B-2 x AAQ48024   ..

 Align seq 1/1  to: AAQ48024  from: 1  to: 997

       18 LeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGl  34
          ::|||'||||       ||||||::||||::
       20 ATTCCACCGGACCCGGCGGAGCCGGCGTCGGGATA.................  52

       34 yAlaValGlyLeuGluArgArgAspTrpSerProSerPro..........  47
          |||:::|||||   :::   |||             |||
       53 .GCCGGTGGGCTCGGGCAGCCCGACGCTCGCTCAGGTTCGCGAGCTTCGGC 101

       48 .....ProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyr  62
            |||      |||      |||
      102 TTGTCCCCCCGCGGCTCCAGTGGGA.......................... 127

       63 PheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGl  79
                          ::|||   :::   |||   ::::||||||
      128 ------------CTCCTCGACGACTCCTCCTGGGCCGCCAAGCCCCCGGT 162

       79 yAlaSerSerArgGluGluProProGluGluProGluGlnCysProVal1  96
          ||||:::   :::   :::|||||   :::   |||          :
      163 GGCCGCGCGCATGGGCAGCCCCGGCTCCGGCGTGAG.............C 200

       96 leAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeu 112
          ::  :::   |||   ||||||||||:|         ||||||||:
      201 TGCCTGCTGTTGCACTTGCTGGTTCTGCCTCCAAGCCCAGGTAACTGT 250

      113 ThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGl 129
          :::
      251 TCA.................................................. 253

      129 uValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaA 146
          ||||||      :::   ||||||||||:::             |||:
      254 .GTCCTCACCTAATTTTACACAGCATGTGAG.............GGAGCAG 290

      146 spProMetAspTrpSerProSerAsnValGln.LysTrpLeuLeuTrpTh 162
          ::|||   |||   |||         ::|||:::   ||||:   ||
      291 AGCCTGGTGACGGATCA.........GCTCAGCCGCCGCCTCATCCGACC 341

      162 rGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGluLeuA 179
          |   :::|||                  ||||||      |||   ||||
      342 C...TACCAGCTCTACAGCCGCACCAGCGGGAAGCACGTGCAGGTCCTGG 378

      179 laGlyLysGluLeuCysAlaMetSerGlnGln 189
          |::::|||:::::   ||||||:::|||:::
      379 CCAACAAGCGCATCAACGCCATGGCAGAAGAC 410

seq_name: US-09-126-945B-2.rng_seq_nos_seq_NA2999.RAT:AAZ21501

seq_documentation_block:
ID   AAZ21501 standard; DNA; 80161 BP.
XX
AC   AAZ21501;
XX
DT   01-DEC-1999  (first entry)
XX
DE   DNA fragment of Saccharopolyspora spinosa containing biosynthetic genes.
XX
KW   Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW   microlides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW   PKS; extender module; initiator module; acyl transferase domain; AT;
KW   acyl carrier protein; ACP; beta ketosynthase domain; KS; KR;
```

```
KW   dehydratase domain; DH; enoyl reductase domain; ER; beta ketoreductase;
KW   insecticide; ss.
XX
OS   Saccharopolyspora spinosa.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1145..1971)
FT                   /*tag   a
FT                   /product  ORFL16
FT                   /note  "Protein involved in transcription control"
FT   CDS             2024..2791
FT                   /*tag   b
FT                   /product  ORFL15
FT                   /note  "Keto acyl reductase"
FT   CDS             complement (3416..4165)
FT                   /*tag   c
FT                   /product  spnG
FT                   /note  "Spinosyn biosynthesis protein G"
FT   CDS             complement (4168..5425)
FT                   /*tag   d
FT                   /product  spnK
FT                   /note  "Spinosyn biosynthesis protein K"
FT   CDS             complement (5363..6751)
FT                   /*tag   e
FT                   /product  spnQ
FT                   /note  "Spinosyn biosynthesis protein Q"
FT   CDS             7083..8450
FT                   /*tag   f
FT                   /product  spnP
FT                   /note  "Spinosyn biosynthesis protein P"
FT   CDS             8967..10427
FT                   /*tag   g
FT                   /product  SpnO
FT                   /note  "Spinosyn biosynthesis protein O"
FT   CDS             10436..11434
FT                   /*tag   h
FT                   /product  spnN
FT                   /note  "Spinosyn biosynthesis protein N"
FT   CDS             complement (11540..12492)
FT                   /*tag   i
FT                   /product  spnM
FT                   /note  "Spinosyn biosynthesis protein M"
FT   CDS             complement (12696..13547)
FT                   /*tag   j
FT                   /product  spnL
FT                   /note  "Spinosyn biosynthesis protein L"
FT   CDS             complement (13592..14785)
FT                   /*tag   k
FT                   /product  spnK
FT                   /note  "Spinosyn biosynthesis protein K"
FT   CDS             complement (14799..16418)
FT                   /*tag   l
FT                   /product  spnJ
FT                   /note  "Spinosyn biosynthesis protein J"
FT   CDS             16556..17743
FT                   /*tag   m
FT                   /product  spnI
FT                   /note  "Spinosyn biosynthesis protein I"
FT   CDS             complement (17749..18501)
FT                   /*tag   n
FT                   /product  spnH
FT                   /note  "Spinosyn biosynthesis protein H"
FT   CDS             complement (18541..19713)
FT                   /*tag   o
FT                   /product  spnG
FT                   /note  "Spinosyn biosynthesis protein G"
FT   CDS             20168..20995
FT                   /*tag   p
FT                   /product  spnF
FT                   /note  "Spinosyn biosynthesis protein F"
FT   CDS             21111..28898
FT                   /*tag   q
FT                   /product  spnA
FT                   /note  "Spinosyn biosynthesis protein A"
FT                   /transl_except  (pos:26940..26942,aa:Pro)
FT   misc_feature    21126..22379
FT                   /*tag   r
FT                   /note  "Beta ketosynthase domain (KSi): part of the
FT                   initiator module"
FT   misc_feature    22692..23869
FT                   /*tag   s
FT                   /note  "Acyl transferase domain (AIi): part of the
FT                   initiator module"
FT   misc_feature    24793..24041
FT                   /*tag   t
FT                   /note  "Acyl carrier protein domain (ACPi): part of the
FT                   initiator module"
FT   misc_feature    24102..25349
FT                   /*tag   u
FT                   /note  "Beta-ketosynthase domain (KS1): part of
FT                   extender module 1"
FT   misc_feature    25683..26684
FT                   /*tag   v
FT                   /note  "Acyl transferase domain (AT1): part of
FT                   extender module 1"
FT   misc_feature    27582..28121
FT                   /*tag   w
FT                   /note  "Beta-ketoreductase domain (ER1): part of
FT                   extender module 1"
FT   misc_feature    28404..28649
FT                   /*tag   x
FT                   /note  "Acyl carrier protein domain (ACP1): part of
FT                   extender module 1"
FT   CDS             28916..35374
FT                   /*tag   y
FT                   /product  spnC
FT                   /note  "Spinosyn biosynthesis protein C"
FT   misc_feature    29024..30295
FT                   /*tag   z
FT                   /note  "Beta ketosynthase domain (KS2): part of
FT                   extender module 2"
FT   misc_feature    30629..31621
FT                   /*tag   aa
FT                   /note  "Acyl transferase domain (AT2): part of
FT                   extender module 2"
FT   misc_feature    31697..32254
FT                   /*tag   ab
FT                   /note  "Dehydratase domain (DH2): part of extender
FT                   module 2"
FT   misc_feature    33035..34072
FT                   /*tag   ac
FT                   /note  "Enoyl reductase domain (ER2): part of
FT                   extender module 2"
FT   misc_feature    34082..34621
FT                   /*tag   ad
FT                   /note  "Beta ketoreductase domain (KR2): part of
FT                   extender module 2"
FT   misc_feature    34886..36295
FT                   /*tag   ae
FT                   /note  "Acyl carrier protein domain (ACP2): part of
FT                   extender module 2"
FT   CDS             35419..44934
FT                   /*tag   af
FT                   /product  spnC
FT                   /note  "Spinosyn biosynthesis protein C"
FT   misc_feature    35518..36786
FT                   /*tag   ag
FT                   /note  "Beta ketosynthase domain (KS3): part of
FT                   extender module 3"
FT   misc_feature    37108..38097
FT                   /*tag   ah
FT                   /note  "Acyl transferase domain (AT3): part of
FT                   extender module 3"
FT   misc_feature    38992..39528
FT                   /*tag   ai
FT                   /note  "Beta ketoreductase domain (KR3): part of
```

```
FT                      extender module 3"
FT   misc_feature    39790..40045
FT                   /*tag= aj
FT                   /note= "Acyl carrier protein domain (ACP3): part of
FT                      extender module 3"
FT   misc_feature    40102..41373
FT                   /*tag= ak
FT                   /note= "Beta ketosynthase domain (KS4): part of
FT                      extender module 4"
FT   misc_feature    41713..42705
FT                   /*tag= al
FT                   /note= "Acyl transferase domain (AT4): part of
FT                      extender module 4"
FT   misc_feature    43615..44157
FT                   /*tag= am
FT                   /note= "Beta-ketoreductase domain (KR4): part of
FT                      extender module 4"
FT   misc_feature    44431..44676
FT                   /*tag= an
FT                   /note= "Acyl carrier protein domain (ACP4): part of
FT                      extender module 4"
FT   CDS             44966..59752
FT                   /*tag= ao
FT                   /product= spnD
FT                   /note= "Spinosyn biosynthesis protein D"
FT   misc_feature    45077..46348
FT                   /*tag= ap
FT                   /note= "Beta ketosynthase domain (KS5): part of
FT                      extender module 5"
FT   misc_feature    46691..47674
FT                   /*tag= aq
FT                   /note= "Acyl transferase domain (AT5): part of
FT                      extender module 5"
FT   misc_feature    47753..48310
FT                   /*tag= ar
FT                   /note= "Dehydratase domain (DH5): part of extender
FT                      module 5"
FT   misc_feature    49226..49771
FT                   /*tag= as
FT                   /note= "Beta ketoreductase domain (KR5): part of
FT                      extender module 5"
FT   misc_feature    50009..50254

alignment_scores:
            Quality:  103.00              Length:     274
              Ratio:  0.811                 Gaps:      16
 Percent Similarity:  46.350   Percent Identity: 25.547

alignment_block
 US-09-126-945B-2 x AAZ21501   ..

 Align seq 1/1  to: AAZ21501  from: 1  to: 80161

        7 GlyLeuSerSerValSerProSerHisLeuLeuLeuProProAspThrVa 23
          |||         ::|||   ::::::|||::|||         :::::
    32477 GGTCGACCGGTCATCTCGCTGGACCGACTCGTATTGCGCTCGGCTGCGCC 32526

       23 lSerArgThrGlyLeuGlnLysAlaAlaAlaGlyAlaValGlyLeuGlnA 19
          :  |||||||||   ::::::  |||     |||     |||   |
    32527 CCGGTGGACCGGTGCGCGCCGACAGGCCCATCAAGCT...CGGTTGTACC 32573

       40 rqArqAspTrp.....................SerProSerProProAla 49
          ||   :::|||                    |||   :::|||||:::
    32574 GGTTGACCTGGCCAACGGTTCAACTGCCGACATCCGCTCAGCCACCGTCC 32623

       50 ThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLe 66
                 |||   |||         :::|||         :::::
    32624 TGCGTGCTTCTCGGCACCTCAGAA......GTGTCCGCTGACATACAGGT 32667

       66 nTyrProGlu...AspSerSerTrpAlaAlaLysAlaProGlyAlaSerS 82
          :|||||||::      |||      ||||||         ||||||
    32668 GTATCCGGACCTCCGGTCGTTGACGGCTGCGTTGGATGCCGGTGCC.... 32713

       82 erArgGluGluProProGluGluProGluGlnCysProValIleAspSer 98
                |||||||||                        ||:::   :::
    32714 .......GAACCACCC.................GGCGTCGTCATCGCA 32757

       99 GlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGl 115
          :::|||   |||                    ||||||   |||   ::
    32738 CCCACGGCCCCCGGC.................GGTGGACGAACAGCGGA 32769

      115 n.............GluHisSerLeuGluGlnValGluSerMetValV 127
          :          :::|||:::|||:::   |||||||::   :::
    32770 TGTGCGGGAGACGACTCGGCATGCACTCGACCTGCTACAAGGCTGGCTTT 32819

      127 alGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIle 143
          :::   |||||:::                :::|||||   :::
    32820 CCGATCAGCGACTCAACGAA.............TCGCGATTGCTCCTGGTG 32857

      144 ThrAlaAspProMetAspTrpSerPro.............SerAsnValGl 156
          |||      :::     ::|||||              ::::::::::
    32858 ACACAGGGAGCAGTGGCCGTGGAGCCGGGCGAACCCGTGACCGATCTGGC 32907

      156 nLysTrpLeuLeuTrp.....................IhrGluHisGlnP 166
           :::      ||||||                    |||||||||
    32908 GCAGGCCGCGCTCTGGGACTGCTGCGGTCGACGCAGACCGAACACCCTG 32957

      166 yrArg...............LeuProProMetGlyLysAlaPheGlnGlu 177
          |||              ::|||     ::::::   :::
    32958 ATCGCTTCGTCCTCGTCGATGTGCCTGAGCCCGCCAACTCCTCCCCGCG 33007

      178 LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArqGln.. 193
          |||   |||          |||:::     :::     :::|||:::
    33008 CTGCCGGGGGTGAGGCTGCGGCGAACCTCAGCTCGCGTTGCGACGTGG 33057

      194 .......ArqSerPro.............LeuGlyGlyAspValLeuHisA 204
              :::::::|||              ||||:::::|||   :::
    33058 CGGCGCTCATGCCGCCCAGACTGGCTGGACTGGGCAGCGATGACGTC.... 33103

      204 laHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArqThrSer 220
            |||   :::          ::::::::|||     |||   |||
    33104 ....CTGCCGTGCCGGACGCACCGGTGGCGATGGAGGCCACGCGC 33148

      221 ProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThr.. 236
          |||||:::::          |||   :::::|||   :::   |||
    33149 CCGGGAAGCCTGGATGGTTGGCATTGGTGGACGAACCGACGGCCACGGC 33198

      237 ..........AspSerGluVal 240
                  |||:::||||||
    33199 ACCGC.GGGTGACGTGAGGTC 33220

seq name   SEQ 9 partial cDNA sequence; cDNA to mRNA. EMBL: AAZ38457

seq_documentation_block:
ID   AAZ38457 standard; cDNA; 4220 BP.
XX
AC   AAZ38457;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human D-SLAM cDNA.
XX
KW   D-SLAM; dendritic enriched, secreted lymphocyte activation molecule;
KW   immunoglobulin superfamily; activation; lymphocyte; differentiation;
KW   proliferation; immune response; interferon gamma; dendritic cell;
KW   antigen presenting cell; diagnosis; treatment; maturation; survival;
KW   haemato-poietic cell; immune disorder; HIV; haemostatic activity;
KW   thrombolytic activity; modulation; autoimmune disorder;
KW   hyperproliferative disorder; cancer; neurodegeneration; nervous system;
KW   stroke; neuropathy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             92..949
FT                   /*tag= a
```

```
FT                   /product   "Human D SLAM"
FT   sig peptide     92..157
FT                   /*tag    b
FT   mat peptide     158..946
FT                   /*tag    c
FT                   /product   "Mature human D SLAM"
XX
PN   WO9940184 A1.
XX
PD   12 AUG 1999.
XX
PF   04 FEB 1999;   99WO US02415.
XX
PR   06 FEB 1998;   98US 0073962.
PR   19 MAR 1998;   98US 0078572.
XX
PA   (HUMA ) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Young P;
XX
DR   WPI; 2000 052627/04.
DR   P PSDB; AAY52479.
XX
PT   New isolated dendritic enriched secreted lymphocyte activation
PT   molecules, used for treating e.g. immune disorders, cancers,
PT   inflammation, transplant rejection or infections, or as food additives
PT
XX
PS   Claim 1; Fig 1; 106pp; English.
XX
CC   This sequence represents cDNA encoding human dendritic enriched secreted
CC   lymphocyte activation molecule (D SLAM). D SLAM is homologous to members
CC   of the SLAM family and thus should have activity similar to other SLAM
CC   family members. SLAM proteins are members of the immunoglobulin
CC   superfamily and are rapidly induced after activation of naive B and
CC   T lymphocytes, causing differentiation and proliferation of immune
CC   cells. To elicit an immune response, both a secreted form of SLAM, in
CC   addition to a membrane bound SLAM, are thought to interact. Binding of
CC   D SLAM induces the production of interferon-gamma from other cell types,
CC   particularly B and T cells. Because of the tissue distribution of
CC   D SLAM, this protein may also play a role in stimulating dendritic or
CC   antigen presenting cells. The polypeptides can be used for preventing,
CC   treating or ameliorating a variety of medical conditions. Pathological
CC   conditions or susceptibility can be diagnosed by determining the
CC   presence or amount of D SLAM polypeptides in a sample or by determining
CC   the presence of mutations in nucleotides encoding D SLAM. D SLAM proteins
CC   can be used to control the proliferation, activation, maturation,
CC   survival and/or differentiation of haematopoietic cells, in particular
CC   B and T cells. They can be used to treat immune disorders which can be
CC   somatic (such as cancer), acquired (e.g., via chemotherapy) or infectious
CC   in origin. Such immune disorders include HIV infection, leukocyte
CC   adhesion deficiency syndrome, phagocyte bactericidal dysfunction or
CC   severe combined immunodeficiency (SCID). D SLAM proteins or nucleotides
CC   may be used to treat deficiencies or disorders of haematopoietic cells
CC   (e.g., anaemia), and to modulate haemostatic or thrombolytic activity.
CC   They may also be used in the diagnosis or treatment of autoimmune
CC   disorders such as rheumatoid arthritis, haemolytic anaemia and type I
CC   diabetes mellitus, and in the treatment of hyperproliferative disorders
CC   such as cancer. They can be used to promote neuroregeneration as a means
CC   of treating central and peripheral nervous system diseases, e.g.,
CC   Alzheimer's disease, Parkinson's disease, neuropathies, or mechanical
CC   and traumatic disorders (e.g., spinal cord disorders, head injury and
CC   stroke).
XX
SQ   Sequence 4220 BP; 832 A; 885 C; 755 G; 748 T; 0 other;


alignment scores:
            Quality:  102.50              Length:     839
              Ratio:  0.748                 Gaps:      19
 Percent Similarity:  40.418   Percent Identity:  23.304

alignment block:
```

```
US-09-126-945B-2 x AAZ48457   ..

Align seq 1/1   to: AAZ48457   from: 1   to: 4220

       2 GlySerAlaSerProGlyLeuSerSerVal..................... 11
         ||||||||||||||||   |||   |||
     199 GGcAGCGcGTCCCCCTGGCTTCCAAGTCCGTGAGGCTATCTGGCGATCTCC 248

      12 ..............SerProSerHisLeuLeu......LeuProProA 21
                       :::|||:::|||::::::      ||||||   |
     249 TCTGGCCTTCAGAAGAGCTCCTGGCCACGTTTTTCCGAGGCTCCCTGGAG 298

      21 spThrVal.......................SerArgThrGlyLeuGlu 29
         |||:|||                       |||
     299 ACTCTGTACCATTCCCGCTTCCTGGGCCGAGCCCAGCTACACAGCAACCT 348

      30 LysAlaAlaAlaGlyAlaValGlyLeuGluArgArg.............. 41
         :::   :::|||   |||:::|||:::   |||:::
     349 CAGCCTGGAGCTCGGGCCGCTGGAGTCTGGAGACAGCGGAACTTCTCCG 398

      42 .........................AspTrpSerProSerProProAlaT 50
                                  :::|||:::|||||||||
     399 TGTTGATGGTGGACACAAGGGGCCAGCCCTGGACCCAGACCCTCCAGCT. 447

      50 hrProGluGluGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeu 66
               ||||||:::   :::
     448 .......CAAGGTGTACCATGC........................... 462

      67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 83
                                :::|||:::|||   |||:::|||   ::
     463 ....................AGTGCCCAGGCCCGTGGTACAAGTGTTCA 491

      83 q.................GluGluProProGluGluProGluGlnCysP 94
         :                :::          :::   |||
     492 TTGCTGTAGAAAGGGATGCTCAGCCCTCCAAGACCTGCCAGGTTTCTTG 541

      94 roValTleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGly 110
         ::::::      |||          :::::|||   |||:::   |||
     542 TCCTGTTGGGCCCCCAACATCAGCGAAATAACCTATAGCTGGCGACGGGA 591

     111 .........GlyLeu.ThrLeuGluGluHisSerLeu.........GlnG 121
                 ||||||    :::|||   ||||||||          |
     592 GACAACCATGGACTTTGGTATGGAACCACACAGCCTCTTCACAGACGGAC 641

     121 InValGluSerMetValValGlyGluValLeuLysAspIleGluThrAla 137
         |||||   |||:::   :::|||          :::|||:::      :::
     642 AGGTGCTGAGCCATTTCCCTGGGACCAGGAGACAGAGATGTGGCCTATTCC 691

     138 CysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAs 154
         |||           |||:::::::::|||:::::|||:::   :::::
     692 TGC...........ATTGTCTCCAACCCTGTCAGCTGGGACTTGGCCAC 729

     154 nValGlnLysTrp.Leu.....LeuTrpThrGluHisGlnTyrArq... 167
         :|||      ||| :::      :::      :::|||||   |||
     730 AGTCACGCCTGGGATAGTGTCATCATGAGGCAGACCAGGGAAGGCCT 779

     167 ................................................. 167

     780 CCTACAAAGATGTGCTGCTGGTGGTGGCCTGTCTCGCTGCTCCTGATG 829

     168 ..............LeuProProMetGlyLysAlaPheGlnGluLeuA 179
                      ||||||      |||          |||
     830 CTGGTTACTCTCTTCTTGCCTGGCACTGGTGCCCCTGCTCAGGGA...A 876

     179 laGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSer 195
           |||:::::::      :::::|||          ||||||||||:::
     877 AAAGAAAAAGGATGTCCATGCTGACAGAGTGGGTCCAGAGACAGAGAACC 926

     196 ProLeu.....................................GlyGl 199
         ||||||                                     |||
     927 CCCTTGTGCAGGATCTGCCATAAAGGACAATATGAACTGATGCCTGGACT 976

     199 yAspVal.....LeuHisAlaHisLeuAspIleTrpLys...SerAlaA 213
```

```
          ;;;;||      ||||||   |||      ;;;|||      ;;;;;;|
      977 ATCAGTAACCCCACTGCACAGGCACACGATGCTCTGGGACATAACTGGTG 1026

      213 IaTrpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSer 229
          |||||   ;;;      ;;;|||          ;;;   |||   ;;;
     1027 CCTGGAAATCACCATGGTCCTCATATCTCCCATGGGAATCCTGTCCTGCC 1076

      230 IhrSerGluGluSerTrpIhrAspSerGluValAspSerSerCysSerGl 246
          ;;;;;;|||;;;                                    ||
     1077 TCGAAGGAGCAG.....................................GTGG 1098

      246 yGluProIleHis 250
          |||||||   |||
     1094 GCAGCCATCACC 1106

seq_name: ;S;;;A;;;Eda;;;Seq;;;p;;Seq;p;;%A;;;;;;A; AAF;;;;

seq_documentation_block:
ID   AAF62743 standard; cDNA; 4220 BP.
XX
AC   AAF62743;
XX
DT   02-MAY-2001  (first entry)
XX
DE   Human D-SLAM cDNA.
XX
KW   Human; dendritic enriched secreted lymphocyte activation molecule;
KW   D-SLAM; cytostatic, antimicrobial, antiparasitic, immunomodulatory,
KW   antiinflammatory; gene therapy; vaccine; cancer; infection;
KW   immunodeficiency; inflammatory disease; lymphadenopathy;
KW   autoimmune disease; graft versus host disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200111046-A1.
XX
PD   15-FEB-2001.
XX
PF   03-AUG-2000; 2000WO-US21130.
XX
PR   05-AUG-1999;   99US-0369248.
PR   17-MAR-2000; 2000US-0190062.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SA,  Young PE;
XX
DR   WPI; 2001-159977/16.
DR   P-PSDB; AAB71852.
XX
PT   Nucleic acids encoding human dendritic enriched secreted lymphocyte
PT   activation polypeptides, useful for preventing, diagnosing and/or
PT   treating tumors, autoimmune diseases and inflammation -
XX
PS   Claim 1; Fig 1; 316pp; English.
XX
CC   The present sequence encodes human dendritic enriched secreted lymphocyte
CC   activation molecule (D-SLAM). D-SLAM polynucleotides and polypeptides
CC   may be used in the prevention, diagnosis and treatment of disorders that
CC   include tumours and metastases, microbial and parasitic infections,
CC   immunodeficiencies, inflammatory diseases, lymphadenopathy, autoimmune
CC   diseases and graft versus host disease. They may also be used to
CC   stimulate peripheral cell tolerance, to destroy some transformed cell
CC   lines, to mediate cell activation, survival and proliferation, to
CC   mediate immune regulation and inflammatory responses and to modulate
CC   immune responses.
XX
SQ   Sequence 4220 BP; 832 A; 885 C; 755 G; 748 T; 0 other;


alignment_scores:
            Quality:  102.50            Length:     339
              Ratio:  0.748              Gaps:      19

Percent Similarity:  40.413   Percent Identity: 23.304

alignment_block:
US-09-126-945B-2 x AAF62743  ..

Align Seq 1/1  to: AAF62743  Frame: 1  To: 4220

       2 GlySerAlaSerProGlyLeuSerSerVal..................... 11
         ||||||||||.|||   |||   |||
     199 GGCAGCGGCGTCCCCCTGGTTCCAAGTCCGTGAGGCTATCTGGGCATCTC 248

      12 ................SerProSerHisLeuLeu.....LeuProProA 21
                         ;;;|||;;;|.|;;;;;;      ||||||   |
     249 TCTGGCCTTCAGAAGAGCTCCTGGCCACGTTTTTCCGAGGCTCCCTGGAG 298

      22 -p??:Val.......................SerArgThrGlyLeuGlu 29
         ||:;;|||                     |||
     299 ACTCTGTACCATTCCCGCTTCCTGGGCCGAGCCCAGCTACACAGCAACCT 348

      30 LysAlaAlaAlaGlyAlaValGlyLeuGluArgArg.............. 41
         ;;;   ;;;|||   |||;;;|||;;;   |||;;;
     349 CAGCCTGGAGCTCGGGCCGCTGGAGTCTGGAGACAGCGGCAACTTCTCCG 398

      42 .........................AspTrpSerProSerProProAlaP 50
                                  ;;;|||;;;|||||||||
     399 TGTTGATGGTGGACACAAGGGGCCAGCCTGGACCCAGACCCTCCAGCT. 447

      50 hrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeu 66
               |||||||;;;   ;;;
     448 ........CAAGGTGTACGATGC........................... 462

      67 FyrProGluAspSerSerIrpAlaAlaLysAlaProGlyAlaSerSerAr 83
                              ;;;|||;;;|||   |||;;;|||   ;;
     463 ......................AGTGCCCAGGCCCGTGGTACAAGTGTTCA 491

      83 q...................GluGluProProGlnGluProGlnGlnCysP 94
         :                ;;;      ;;;   |||
     492 TTGCTGTAGAAAGGGATGCTCAGCCCTCCAAGACCTGCCAGGTTTTCTTG 541

      94 roValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGly 110
         ;;;;;;      |||          ;;;;;;;||   ||||;;   |||
     542 TCCTGTTGGGCCCCCAACATCAGCGAAATAACCTATAGCTGGCGACGGGA 591

     111 .........GlyLeu.IhrLeuGluGlnHisSerLeu.........GluG 121
                 ||||||    ;;;|||   |||||||||           |
     592 GACAACCATGGACTTTGGTATGGAACTACACAGCCCTTCACAGACGGAC 641

     121 lnValGlnSerMetValValGlyGluValLeuLysAspIleGluIhrAla 137
         |||||   |||;;;   ;;;|||          ;;;|||;;;      ;;;
     642 AGGTGCTGAGCATTTCCCTGGGACCAGGAGACAGAGATGTGGCCTATTCC 691

     138 CysLysLeuLeuAsnIleIhrAlaAspProMetAspIrpSerProSerAs 154
         |||           |||;;;;;;;;;;|||;;;;;;|||;;;   ;;;;;
     692 TGC...........ATTGTCTCCAACCCTGTCAGCTGGGACTTGGCCAC 729

     154 nValGlnLysIrp.Leu.....LeuIrpIhrGluHisGlnTyrArq... 167
         :|||      ||| ;;;      ;;;      ;;;||||||   |||
     730 AGTCACGGCCCTGGGATAGCTGTCATCATGAGGCAGCACCACGGGAAGGCCT 779

     167 ..................................................... 167

     780 CCTACAAAGATGTGCTGCTGGTGGTCGTGCCTGTCCGCTGCTCCTGATG 829

     168 ................LeuProProMetGlyLysAlaPheGlnGluLeuA 179
                         ||||||      |||          |||
     830 CTGGTTACTCTCTTCTGTGCTGGCACTGGTGGCCCCTGCTCAGGGA...A 876

     179 laGlyLysGluLeuLysAlaMetSerGluGluGlnPheArgGlnArgSer 195
           ||;;;;;;;      ;;;;;|||          ||||||||||;;;
     877 AAAGAAAAAGGATGTCCATGCTGACAGAGTGGGTCCAGAGACAGAGAACC 926

     196 ProLeu.....................................GlyGl 199
         ||||||                                     |||
```

```
      927 CCCTTGTGCAGGATCTGCCATAAAGGACAATATGAACTGATGCCTGGACT 976

      199 yAspVal.....LeuHisAlaHisLeuAspIleTrpLys...SerAlaA 213
            :::|||      ||||||   |||      :::|||      :::::|
      977 ATCACTAACCCCACTGCACAGGCACACGATGCTCTGGGACATAACTGGTG 1026

      214 laTrpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSer 229
          |||||   :::      :::|||            :::   |||   :::
     1027 CCTGGAAAICACCATGTTCCICATATICICCCAIGGGAACCCTGTCCTGCC 1076

      230 ThrSerGluGluSerIrpThrAspSerGluValAspSerSerCysSerGl 246
          ::::::|||:::                                    ||
     1077 TCGAAGGAGCAG.............................CCTGG 1093

      246 yGlnProIleHis 250
          |||||||   |||
     1094 GCAGCCAICACAC 1106

Seq name, /SIS3/ftp/data/geneseq/pro_seqs_NA:GENESEQ_AA.fasta

Seq documentation block:
ID   AAZ19894 standard; cDNA; 1078 BP.
XX
AC   AAZ19894;
XX
DT   06-DEC-1999  (first entry)
XX
DE   Human adult blood secreted protein q21 1 cDNA.
XX
KW   Secreted protein; q21 1; human; therapy; diagnosis; vaccine;
KW   blood; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             10..720
FT                   /*tag   a
FT                   /note  "this region is specifically claimed in
FT                    Claim 12(b)"
FT   Sig peptide     10..246
FT                   /*tag   b
FT   mat peptide     247..717
FT                   /*tag   c
FT                   /note  "this region (plus the stop codon) is
FT                    specifically claimed in Claim 12(c) "
FT   sig peptide     10..102
FT                   /*tag   d
FT                   /note  "alternative signal sequence"
FT   mat peptide     103..717
FT                   /*tag   e
FT                   /note  "alternative mature protein"
XX
PN   WO9947555-A1.
XX
PD   23-SEP-1999.
XX
PF   18-MAR-1999;   99WO-US05949.
XX
PR   20-MAR-1998;   98US-0078803.
PR   17-MAR-1999;   99US-0078803.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins Kacie LA,  Evans C;
PI   Merberg D,  Treacy M,  Agostino MJ,  Steininger RJ;
XX
DR   WPI; 1999-562059/47.
DR   P-PSDB; AAY41829.
XX
PT   New polynucleotides derived from murine foetal cell cDNA libraries,
PT   potentially used as, e.g. vaccines
XX
PS   Claim 12(a); Page 93; 107pp; English.
XX
CC   This is the nucleotide sequence of an isolated cDNA clone that
CC   codes for a novel human secreted protein, q21 1 (see AAY41829).
CC   The clone was isolated from a human adult blood (peripheral blood
CC   mononuclear cells treated with phytohaemagglutinin, phorbol
CC   myristate acetate and mixed lymphocyte reaction) cDNA library
CC   using human cDNAs that had themselves been identified in a human
CC   adult blood cDNA library using methods selective for cDNAs encoding
CC   secreted proteins.  The clone is deposited in composite deposit
CC   ATCC 98700.  The invention provides new human secreted proteins
CC   (see AAY41828-38) and polynucleotides (see AAZ19893-901) isolated from
CC   foetal cell, adult blood, adult brain and foetal kidney cDNA
CC   libraries.  They are predicted to have biological activities which
CC   would make them suitable for treating, preventing or ameliorating
CC   medical conditions in humans and animals, although no supporting
CC   data are given.  Suggested activities include nutritional,
CC   cytokine, tissue growth, cell proliferation and differentiation,
CC   immunostimulant (e.g. as vaccine), immunosuppressive,
CC   tumour protein regulating, activin or inhibin, chemotactic or
CC   chemokinetic, haemostatic or thrombolytic, receptor/ligand activity,
CC   antiinflammatory, cadherin or tumour invasion suppressor, and
CC   tumour inhibition activities.  The polynucleotides are also stated
CC   to be useful for gene therapy.
XX
SQ   Sequence 1078 BP; 244 A; 311 C; 332 G; 199 T; 2 other;


alignment scores:
            Quality:  102.000              Length:    258
              Ratio:  0.887                  Gaps:     16
 Percent Similarity:  44.574   Percent Identity: 24.031

alignment block:
 US-09-126-945B-2 x AAZ19894   ..

 Align seq 1/1  to: AAZ19894  from: 1  to: 1078

       15 HisLeuLeuLeuProProAspIhrValSerArqThrGlyLeuGluLysAl 41
          |||:||:::|||      :::|||:::   |||         :::   ||
      205 CACCTGCTGCTGCTGCCTGAACACTCTGCGGCGCATCCAGGCGTCCATGGC 254

       41 aAlaAlaGlyAlaValGlyLeuGluArqArqAspTrpSerProSerProP 48
          |   |||:::|||:::                              ||||||||||
      255 ACCCCGGCTGCCCTG..............CCACCTGTGCCTAGCCCAC 289

       48 roAlaIhrPro.............................GluGlnGly 54
          |||||::|||                               :::   :::
      290 CTGCAGCCCCCAGTGTGGCTGACAACTTACTGCCAAGCTCGGACGCTGCC 339

       55 LeuSerAlaPheTyrLeuSerTyrPhe...AspMetLeuTyrProGluAs 70
          |||||||||         |||   :::   ||||:   :::   ||
      340 CTTTCAGCCTCCATGGCCAGCCTCCTGGAGGACCTCAGCCACATTGAGGG 389

       70 pSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArqGluGluProP 87
             |||         :::|||||||         :::   |||   ||||
      390 CCTGAGT.........CAGGCTCCCCAACCCTTGGCAGASGAGGGGCCAC 430

       87 roGluGloProGluGlnCysProValIleAspSerGluAlaProAla... 102
          ||   :::                  |||   :::   |||||||:::
      431 CAGGCCGIAGC..............AICGGGGGAGCAGCGCCCAGCCIG 465

      103 GlySerLeuAspLeuVal...ProGlyGlyLeuThrLeuGluGluHisSe 118
          |||:::|||||||||:::   |||:::         |||   :::   ::
      466 GGAGCCTIGGACCTGCTGGGCCAGCCACTGGCTGTCTACTGGACGAGAGG 515

      118 rIeuGluGluValGlnSerMetValValGlyGluValLeuLysAspIleG 135
          :|||||                :::::::::::::::
      516 GCTTGAG.....................GGCCTGTTGAGGATATTG 541

      135 luThrAlaCysLysLeuLeuAsnIleThrAlaAspProMet..... 148
          ::|||:: all      :::|||:::      :::|||:::
      542 ACACCICIAIGIAIGACAAIGAACTTIGGGACCAGCCICIGAGGGCIC 591
```

```
     149 ..............................................AspIrpS 151
                                                      :::::::
     592 AAACCAGGCCCTGAGGATGGGCCGGGCAAGGAGGAAGCTCCGGAGCTGGA 641

     151 erIrpSerAsnValGlnLysIrpLeuLeuIrpThrGluHisGlnIyr 166
         ::  :|||::  :::  :::|||::  |||     |||::::::
     642 CGAGGCGAATTGGACTACCTCATGGATGTGCTGGTGGGCACACAGGCAC 691

     167 ArqLeuProProMetGlyLysAlaPheGlnGlu..................... 177
                    |||::::::  :::
     692 TGCAGCCACCGCGGGCGCCAGGGCGCTGAGCCCTCCTGCTGCAATGCTTC 741

     178 ........................LeuAlaGlyLysGluLeuCysAlaMetS 187
                               |||||||||     :::     :
     742 TCTGGTATCTGAACTGAGCCTGCTGGCTGGACCAACTGTCCTCCAAAAGA 791

     187 erGlu...GluGlnPheArqGlnArqSerProLeuGlyIyAspValLeu 232
         :::::     ::::::  |||||||::  ||||||
     792 CACAGCTGGCTTCCCTAGTACAGAGAACAGGGCTTGGG.............. 829

     233 HisAlaHisLeuAspIleIrpLysSerAlaAlaIrpMetLysGlnArq.I 279
                 ::|||                  :::::::|| :
     830 .............CCACTTTGG.................AGAGACAGAAI 849

     279 hrSerProGlyAlaIleHis 225
         ::|||||||||   :::|||
     850 CTAGTCCTGGGCAACTTCAC 869
```

```
Seq_name: ????????????????????pMON42272.seq;
```

```
seq documentation_block:
ID   AAV55466 standard; DNA; 1089 BP.
XX
AC   AAV55466;
XX
DT   25-NOV-1998  (first entry)
XX
DE   Flt3L/IgG2b/G-CSF chimeric pMON42272.seq;
XX
KW   Haematopoietic receptor agonist; human; flt-3 receptor agonist;
KW   flt-3 ligand; Flt3L; granulocyte colony stimulating factor; G-CSF;
KW   chimeric protein; stem cell expansion; tumour; infection;
KW   autoimmune disease; haematopoietic disorder; therapy;
KW   dendritic cell; pMON42272; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9817810-A2.
XX
PD   30-APR-1998.
XX
PF   24-OCT-1997;   97WO-US020037.
XX
PR   25-OCT-1996;   96US-0029629.
XX
PA   (SEAR ) SEARLE & CO G D.
XX
PI   Feng Y, McKearn JP, McWherter CA, Minnerly JC, Minster NI;
PI   Staten NR, Streeter PR, Summers NL, Woulfe SL;
XX
DR   WPI; 1998-261504/23.
DR   P-PSDB; AAW78074.
XX
PT   Multi-functional chimeric haematopoietic receptor agonist useful
PT   to treat haematopoietic disorders, tumours, infections or autoimmune
PT   diseases
XX
PS   Example 110; Page 167; 841pp; English.
XX
CC   This nucleic acid molecule, designated pMON42272.seq, codes for a
CC   dimeric polypeptide (see AAW78074) consisting of a sequence
CC   rearranged flt3 ligand (see AAW77782) joined via an IgG2b linker to
CC   granulocyte colony stimulating factor (G-CSF) (see AAW77784). Novel
CC   multi-functional chimeric haematopoietic receptor agonists (see e.g.
CC   AAW77812-22) can include these chimeric G-CSF and flt-3 receptor
CC   agonists. They are used to stimulate the production of haematopoietic
CC   cells in a patient, for the ex vivo expansion of haematopoietic cells,
CC   for the production of dendritic cells and to treat haematopoietic
CC   disorders, tumours, infection or autoimmune diseases. Nucleic acid
CC   molecules encoding the receptor agonists can be used for the gene
CC   therapy of haematopoietic disorders.
XX
SQ   Sequence 1089 BP; 192 A; 357 C; 399 G; 214 T; 0 other;
```

```
alignment_scores:
          Quality:  102.00          Length:    216
            Ratio:  1.030             Gaps:     11
 Percent Similarity:  45.833   Percent Identity: 26.389

alignment_block:
 US-09-126-945B-2 x AAV55466    ..

 Align seq 1/1  to: AAV55466  from: 1  to: 1089

       8 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
         ||||||   |||   ||.|||||||::   ||||||||  |||   ||
     147 TCAGCCCCCCCCCAGCTGTCTTCGCTCGTCCAGACCAACATCTCCCGCC 196

      19 oProAspThrValSerArqThrGlyLeuGluLysAlaAlaAlaGlyAlaV 36
         ||||             ::|||      :::||||||||:.||||
     197 TCCT..............GCAGGAGACCTCCGAGCAGCTGGTGGCGCT. 230

      36 alGlyLeuGluArqArqAspIrpSerProSerProProAlaThrProGlu 52
               |||      |||   ||:|||||:          ::::||
     231 ........GAAGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGGA 272

      53 GluGlyLeuSerAlaPhelyrLeuSerTyrPheAspMetLeuTyrProGl 69
         ::::::?||||     ||||::
     273 GCTGCAGTGTCAGCCGACTCCTCAAC..................... 299

      69 nAspSerSer|;AlaAlaLys.AlaProGlyAlaSerSer.ArqGlnGl 85
         :::   |:|:::   ::: :::  ||||||||||||    ||:::
     300 .CCTGGGCGGTGGGTCAGGAGGTGGGTCAGGCGGTGGTTCTGGCGGAGGA 348

      85 nProProGluGluProGluGlnCys.........ProValIleAspSerG 99
         :|||      |||   :||   :::           |:|:::
     349 TCCGGCGGCCGAAGCGGAGTGGCTCTGGGGGAGGTAGTGGTACCCAGGA 398

      99 InAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGlu 115
         ||||||:::::::::          |||   :::||:::
     399 CTGCTCCTICCAACACAGCCCCATCTCCTCCGACTTCGCTGTCA..... 442

     116 GluHisSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLy 132
                                                        ||
     443 ..................................................AA 444

     132 sAspIleGluLrAlaCysLysLeuLeuAsnIleThrAlaAspProMetA 147
         |:::::::::                 ::::::||||||   ::::||
     445 ArCCAGTGAGTGTCTGACTACCTGCTTCAAGATTACCCAGTCACCG.... 490

     149 spIrpSerProSer.AsnValGluLys.....................Ir 158
           |||   ||:::::    |||::::::                   ||
     491 ..TGGCCTCCAACCTGCAGTACGTAGAGGGCGGTGGAGGCTCCCCGGGTG 538

     158 pLeuLeuIrp......IhrGluHisGlnIyrArqLeuProProMetGlyL 173
         |   ::::|i         ::|||   |||   :::   |||
     539 GTGGTTCTGGCGGCGGCTCCAACATGGCTACACCATTGGGCCCTGCCAGC 588

     173 ysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGlu 189
                 ||||||   ::::::   :::   |||   ::::||:::
     589 TCCCTGCCCCAGAGCTTCCTGCTCAAGTCTTTAGAGCAAGTGAGGAAGAT 638

     190 GlnPheArqGlnArqSerProLeuGlyGlyAspValLeuHis 204
               |||   ||:|||      |||      |||   ||
```

```
      639 CCAGGGCGATGGCTCAGCGCTCCAGGAGAAGCTGTGTGCCAC 680

seq name: /cgi6/ptodata/patseq/genseqp/NA1990.DAT:AAQ05686

seq documentation block:
ID   AAQ05686 standard; DNA; 3243 BP.
XX
AC   AAQ05686;
XX
DT   03-JAN-1991  (first entry)
XX
DE   Islets of Langerhans cell clone ICA12.3 (ATCC 40703).
XX
KW   Pancreatic Islet Cell Antigen (ICA) 12.3; ATCC 40703;
KW   insulin dependent (Type I) diabetes mellitus; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             692..1618
FT                   /*tag= a
XX
PN   EP383129-A.
XX
PD   22-AUG-1990.
XX
PF   03-FEB-1990;   90EP-0102179.
XX
PR   04-DEC-1989;   89US-0441703.
PR   17-FEB-1989;   89US-0312543.
XX
PA   (MOLE-) MOLECULAR DIAGNOSTIC.
XX
PI   Rabin D;
XX
DR   WPI; 1990-255450/34.
DR   P-PSDB; AAR06467.
XX
PT   Pancreatic islet cell antigens obtd. by mol. cloning - used in
PT   research, diagnosis and treatment of insulin-dependent (type I)
PT   diabetes mellitus.
XX
PS   Disclosure; Fig 14; 45pp; English.
XX
CC   A cDNA library was constructed from RNA extracted from purified
CC   human islets of Langerhans cells. The library was screened with
CC   sera from autoimmune patients with Type I diabetes mellitus.
CC   ICA 12.3 is one of nine clones found to react predominantly with
CC   diabetic sera. The protein expressed by the clone was analysed by
CC   expressing the clones in E coli hosts. The protein or peptide
CC   fragments of it can be used as immunoassay reagents in the
CC   prosymptomatic diagnosis of insulin dependent diabetes mellitus
CC   (IDDM).
CC   They may also be used to identify or mark T- or B-cells involved in
CC   IDDM or to bind or block these same cells for therapy.
CC   See also AAQ05681 Q05685 and AAQ05687 Q05688.
XX
SQ   Sequence 3243 BP; 664 A; 1007 C; 853 G; 719 T; 0 other;


alignment scores:
            Quality:  101.50              Length:     134
              Ratio:  1.346                 Gaps:       8
 Percent Similarity:  56.716   Percent Identity: 27.612

alignment block:
 US-09-126-945B-2 x AAQ05686/rc ?  ..

 Align seq 1/1  to reverse of: AAQ05686  from: 1  to: 3243

      209 IrpLysSerAlaAlaIrpMetLysGluArgThrSerProGlyAlaIle.. 224
          |||  :::|||:::      ::::::  |||||||||||||||
     1241 TGGGCCGCGGCTTGTACTTGTAGTCAGGATACTTCTCCAGGTGCTGCCGG 1192

      225 ....................HistyrCysAlaSerThrSerGluGluSerT 235
                               |||      |||:::::::          ||||
     1191 CTCAGCCGCGCCGTGTTCCTCATAGTAGGGCTGCTTCTCCTGGTGGTCAT 1142

      235 rpThrAspSerGluValAspSerSerCys.SerGlyGlnProIleHisIle 251
          |||||:::|||||||:::::::::|||||| ::::::      :::|||::
     1141 GGACTTCCAGCGAGATCCAAGGATCTGCTGATGCTGGAGTTGTGCATGT 1092

      251 trIrpGlnPheLeuLysGlu.........LeuLeuLeuLysProHisSerT 265
          :|||:::  |||::::::          |||:::|||  ||||||
     1091 CTGGGAAGGCTTGTAGGATCTTCTCCAGCTCATCCTTGGCCACACCATG 1042

      265 yrGly......Arg.PheIleArgTrp...LeuAsnLysGluLysGlyIl 278
          |||          ::::::  |||  |||  ::::::::::|||
     1041 AAGGCGTTCATGGGCCTCTTGATGGGCTGCTGTTTCGGGACTCGGGGAA 992

      278 ePheLysIleGluAspSerAlaGluValAlaArgLeuTrpGlyIleArgL 295
          ::::::     :::  :::|||::::::      :::
      991 GTGGCGGGAGCCATCCATGTCGCAGCTCAGCATGGCTTCCTCCAGTGGGT 942

      295 ysAsnArgPro.................AlaMetAsnTyrAspLysLeu 305
          :::::::||                 |||  :::  ::::::|||
      941 CCACACAGCCGTCCTCCAGCCGCTCCTTGCTGGGAICAGTCCAGCTG 892

      306 SerArgSerIleArgGluTyrTyrLysLysGlyIleIleArgLysPro 321
          ||||||:::|||          ||||||::::::   :::|||
      891 ATGAGGTCCTIACGG............AACGGGGTCTTGGGGGAGCCA 856

seq name: /cgi6/ptodata/patseq/genseqp/NA1990.DAT:AAZ32338

seq documentation block:
ID   AAZ32338 standard; cDNA; 3243 BP.
XX
AC   AAZ32338;
XX
DT   21-JAN-2000  (first entry)
XX
DE   Human pancreatic islet cell antigen ICA 12.3 ATCC 40703 cDNA sequence.
XX
KW   Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
KW   insulin dependent type I diabetes mellitus; diagnosis; detection;
KW   immunoglobulin; T-cell; B-cell; antibody binding; ss.
XX
OS   Homo sapiens.
XX
PN   US5981700-A.
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0468579.
XX
PR   05-MAY-1994;   94US-0239276.
PR   14-JUN-1991;   91US-0715181.
PR   08-JUN-1992;   92US-0872646.
PR   17-FEB-1989;   89US-0312543.
PR   04-DEC-1989;   89US-0441703.
XX
PA   (FARB-) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
XX
PI   Rabin DU;
XX
DR   WPI; 1999-633320/54.
DR   P-PSDB; AAY40852.
XX
PT   New pancreatic islet cell antigens, useful in the diagnosis of
PT   insulin dependent (type I) diabetes mellitus
XX
PS   Claim 1; Column 45-42; 61pp; English.
XX
CC   The present sequence encodes a human pancreatic islet cell antigen (ICA)
CC   that binds with antibodies found in the sera of patients afflicted with
CC   insulin dependent (type I) diabetes mellitus (IDDM). ICAs from the
```

```
CC   present invention are encoded by the DNA insert of a recombinant cloning
CC   vehicle selected from ATCC 40552 (AAZ32337, encoding AAY49847), ATCC
CC   40553 (AAZ32344, encoding AAY49848), ATCC 40554 (AAZ32335, encoding
CC   AAY49849), ATCC 40551 (AAZ32336, encoding AAY49850), ATCC 40552
CC   (AAZ32337, encoding AAY49851), ATCC 40703 (AAZ32338, encoding AAY49852),
CC   ATCC 40704 (AAZ32339, encoding AAY49853), ATCC 40705 (AAZ32340), ATCC
CC   40706 (AAZ32341, encoding AAY49854) and ATCC 75630 (AAZ32342, encoding
CC   AAY49855). ICA proteins and their peptide fragments can be used in the
CC   diagnosis of IDDM and in detecting or blocking human immunoglobulin,
CC   T-cells or B-cells involved in IDDM.
XX
SQ   Sequence 3243 BP; 664 A; 1007 C; 853 G; 719 T; 0 other;


alignment_scores:
            Quality:  101.50              Length:    134
              Ratio:  1.336                 Gaps:      8
 Percent Similarity:  56.716   Percent Identity: 27.612

alignment_block:
 US-09-126-945B-2.rx AAZ32338/rev   ..

 Align seq 1/1  to reverse of: AAZ32338  from: 1  to: 3243

     209 IrpLysSerAlaAlaIrpMetLysGluArgThrSerProGlyAlaIle.. 224
         |||   :::|||:::      ::::::   |||||||||||||||
    1241 TGGGCCGCGGCTTGTACTTGTAGTCAGGATACTTCTCCAGGTGCTGCCGC 1192

     225 ...................HisTyrCysAlaSerThrSerGluGluSer 249
                             |||      |||::::::         |||
    1191 CTCAGCCGCGCCTGTTCCTCATAGTAGGGCTGCTTCTCCTGGTTGCTCAT 1142

     250 rpThrAspSerGluValAspSerSerCys.SerGlyGlnProIleHisLe 293
         ||||:::|||||:::::::::::|||||| ::::::     :::||||::
    1141 GGACTTCCAGCCACATCCAAGCATCTTGCTGATGCTGGAGTTGTCCATGT 1092

     294 aTrpGlnPheLeuLysGlu.....LeuLeuLeuLysProHisSerT 265
         :|||:::  |||::::::         |||:::|||  ||||||
    1091 CTGGGAAGGCTTGCAGGATCTTCCTCCGCTCATCCTTGGCCCACACCATG 1042

     265 yrGly.....ArgPheIleArgTrp...LeuAsnLysGluLysGlyIl 278
          |||          ::::::   |||  |||  ::::::::::|||
    1041 AAGGCGTTCATGGGCCTCTTGATGTGGCTGCTGTTCGGGGACTCGGGCAA 992

     278 ePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgL 295
         ::::::      :::  :::|||:::::      :::
     991 GTGGCGGCAGCCATCCATGTCGCAGCTCAGCATGGCTTCCTCCAGTGGGT 942

     295 ysAsnArgPro.................AlaMetAsnTyrAspLysLeu 305
         :::::|||                 |||  :::  ::::::|||
     941 GCACACAGCCGTCCTCCAGCCGCTCCTTGGCTGTGGCATCAGTCCACGCTG 892

     306 SerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
         ||||||::|||            ||||||::::::   ::|||
     891 ATGAGGTCCTTACGG............AAGGGGGTGTTGGGGGAGCCA 856

Seq_name: /cs_sas/1_update/genseq1pro_sep_NA1999_DAT:AAZ09919

seq_documentation_block:
ID   AAZ09919 standard; cDNA; 3243 BP.
XX
AC   AAZ09919;
XX
DT   29-NOV-1999  (first entry)
XX
DE   Human islet cell antigen clone ICA-12.3 cDNA.
XX
KW   Islet cell antigen; ICA; human; pancreas; antibody; IDDM; immunoassay;
KW   insulin-dependent type I diabetes mellitus; diagnosis; immunoglobulin;
KW   T-cell; B-cell; islet cell destruction; immunoreactivity; disease;
KW   immune therapy; stimulating antigen; screening; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             692..1618
FT                   /*tag= a
FT                   /product= "ICA-12.3"
XX
PN   US5955345-A.
XX
PD   21-SEP-1999.
XX
PF   06-JUN-1995;    95US-0468576.
XX
PR   05-MAY-1994;    94US-0239276.
PR   14-JUN-1991;    91US-0715181.
PR   08-JUN-1992;    92US-0872646.
PR   17-FEB-1989;    89US-0312543.
PR   04-DEC-1989;    89US-0441203.
XX
PA   (FARB ) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
XX
PI   Rabin DU;
XX
DR   WPI; 1999-568404/47.
DR   P-PSDB; AAY83358.
XX
PT   Isolated DNA encoding antigens reactive with serum from diabetics
PT   useful to develop products for the diagnosis and monitoring of
PT   insulin-dependent diabetes mellitus
XX
PS   Claim 1; Column 37-44; 64pp; English.
XX
CC   This invention describes novel isolated and purified DNAs encoding
CC   pancreatic islet cell antigens reactive with islet cell antibodies
CC   found in sera of patients with insulin-dependent type I diabetes
CC   mellitus (IDDM). The polypeptides of the invention are useful as
CC   immunoassay reagents in the presymptomatic diagnosis of
CC   insulin dependent type I diabetes mellitus (IDDM). The ability of the
CC   polypeptides to bind the antibody binding site on islet cell antibodies
CC   (ICAs) also confers utility in the binding or blocking of human
CC   immunoglobulin, T-cells, or B-cells involved in IDDM. The products can
CC   also be used to study the biological mechanisms involved in islet cell
CC   destruction and the appearance of ICA. The immunoreactivity profile with
CC   different antigens can provide diagnostically significant information
CC   concerning the nature of the disease, e.g. subtypes, the state of the
CC   disease, the proximity to onset of the disease, and the efficacy of
CC   therapy, e.g. immune therapy. The antigens can serve as stimulating
CC   antigens for T-cell culture, permitting significantly improved T cell
CC   cloning, identification, and growth. The availability of large quantities
CC   of pure antigen enables the development of highly sensitive and specific
CC   immunoassays which can be used to screen the general population of
CC   presymptomatic IDDM or a predisposition to develop IDDM. This sequence
CC   encodes the human islet cell antigen from clone ICA-12.3.
XX
SQ   Sequence 3243 BP; 664 A; 1007 C; 853 G; 719 T; 0 other;


alignment_scores:
            Quality:  101.50              Length:    134
              Ratio:  1.336                 Gaps:      8
 Percent Similarity:  56.716   Percent Identity: 27.612

alignment_block:
 US-09-126-945B-2.rx AAZ09919/rev   ..

 Align seq 1/1  to reverse of: AAZ09919  from: 1  to: 3243

     209 IrpLysSerAlaAlaIrpMetLysGluArgThrSerProGlyAlaIle.. 224
         |||   :::|||:::      ::::::   |||||||||||||||
    1241 TGGGCCGCGGCTTGTACTTGTAGTCAGGATACTTCTCCAGGTGCTGCCGC 1192

     225 ...................HisTyrCysAlaSerThrSerGluGluSer 249
                             |||      |||::::::         |||
    1191 CTCAGCCGCGCCTGTTCCTCATAGTAGGGCTGCTTCTCCTGGTTGCTCAT 1142
```

```
      235 rpThrAspSerGluValAspSerSerCys.SerGlyGlnProIleHisLe 251
          ||||||:||||||:::::::::|||||| ::::::      :::|||::
     1141 GGACTTCCAGCGAGATCCAAGGATCTTGCTGATGCTGGAGTTGTGCATGT 1092

      251 rTrpGlnPheLeuLysGlu.....LeuLeuLeuLysProHisSerI 265
          :|||:::   |||:::::::         |||::|||   ||||||
     1091 CTGGGAAGGCTTGCAGGATCTTCCTCCGCTCATCCTTGGCCCACACCATG 1042

      265 yrGly.....Arg.PheIleArgTrp...LeuAsnLysGluLysGlyIl 278
          |||          ::::::   |||   |||   :::::::::|||
     1041 AAGGCGTTCATGGGCCTCTTGATGTGGCTGCTGTTCGGGACTCGGGGAA 992

      278 rPheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgL 295
          :::::::      :::   :::|||::::::         :::
      991 GTGGCGGGAGCCATCCATGTCGCAGCTCAGCATGGCTTCCTCCAGTGGGT 942

      295 ysAsnArgPro.....AlaMetAsnTyrAspLysLeu 305
          :::::|||          |||   :::   :::::|||
      941 GCACACAGCCGTCCTCCAGCCGCTCCTTGGCTGGGGATGAGTCCAGGCTG 892

      306 SerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
          |||||||::|||           |||||||::::   :::|||
      891 ATGAGGTCCTTACGG...........AAGGGGCTGTTGGGGGAGCCA 856

Seq name: /SIDSB/pubdata/genseqp/genseqp/NA1999 :AT:AAV63557

Seq documentation block:
ID   AAV63557 standard; cDNA; 4243 BP.
XX
AC   AAV63557;
XX
DT   01-FEB-1999 (first entry)
XX
DE   Islet cell antibody antigen cDNA from clone ICA 12.3.
XX
KW   Pancreatic islet cell antibody antigen; ICA antigen; diagnosis;
KW   treatment; insulin dependent diabetes mellitus; IDDM; autoantibody;
KW   autoimmune disease; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             692..1618
FT                   /*tag = a
XX
PN   US5840846-A.
XX
PD   24-NOV-1998.
XX
PF   05-MAY-1994;    94US-0239276.
XX
PR   14-JUN-1991;    91US-0715181.
PR   17-FEB-1989;    89US-0312543.
PR   04-DEC-1989;    89US-0441704.
PR   08-JUN-1992;    92US-0872646.
PR   05-MAY-1994;    94US-0239276.
XX
PA   (FARB ) BAYER CORP.
XX
PT   Rabin DU;
XX
DR   WPI; 1999-044118/03.
DR   P-PSDB; AAW80484.
XX
PT   Polypeptide antigens from pancreatic islet cells - useful for the
PT   detection and treatment of insulin-dependent diabetes mellitus
XX
PS   Example 1; Columns 35-42; 47pp; English.
XX
CC   The present sequence encodes a pancreatic islet cell antibody
CC   (ICA) antigen. The proteins are used for the diagnosis and treatment
CC   of insulin dependent (Type I) diabetes mellitus (IDDM). They are
CC   recognised by autoantibodies of the disease as IDDM is now thought
CC   to be an autoimmune disease. They were identified from probing of
CC   IDDM serum with human cDNA libraries expressing protein fragments.
XX
SQ   Sequence 4243 BP; 664 A; 1007 C; 854 G; 719 T; 0 other;


Alignment scores:
            Quality:  101.50              Length:     134
              Ratio:  1.346                 Gaps:       8
 Percent Similarity:  56.716   Percent Identity: 27.612

Alignment block:
 US-09-126-945B-2 x AAV63557/rev    ..

 Align seq 1/1  to reverse of: AAV63557  from: 1  to: 4243

      209 TrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIle.. 224
          |||   :::|||:::       ::::::   |||||||||||||||
     1241 TGGGCCGCGCCTTGTACTTGTAGTCAGGATACTTCTCCAGCTGCTGCCGG 1192

      225 ....................HisTyrCysAlaSerThrSerGlaGluSerI 245
                               |||      |||:::::          ||||
     1191 CTCAGCCGCGCCCTGTTCCTCATAGTAGGGCTGCTTCTCCTGGTTGGTCAT 1142

      246 rpThrAspSerGluValAspSerSerCys.SerGlyGlnProIleHisLe 251
          |||||:::||||||:::::::::|||||| ::::::      :::|||::
     1141 GGACTTCCAGCGAGATCCAAGGATCTTGCTGATGCTGGAGTTGTGCATGT 1092

      251 rTrpGlnPheLeuLysGlu.....LeuLeuLeuLysProHisSerI 265
          :|||:::   |||:::::::         |||::|||   ||||||
     1091 CTGGGAAGGCTTGCAGGATCTTCCTCCGCTCATCCTTGGCCCACACCATG 1042

      265 yrGly.....Arg.PheIleArgTrp...LeuAsnLysGluLysGlyIl 278
          |||          ::::::   |||   |||   :::::::::|||
     1041 AAGGCGTTCATGGGCCTCTTGATGTGGCTGCTGTTCGGGACTCGGGGAA 992

      278 rPheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgL 295
          :::::::      :::   :::|||::::::         :::
      991 GTGGCGGGAGCCATCCATGTCGCAGCTCAGCATGGCTTCCTCCAGTGGGT 942

      295 ysAsnArgPro.....AlaMetAsnTyrAspLysLeu 305
          :::::|||          |||   :::   :::::|||
      941 GCACACAGCCGTCCTCCAGCCGCTCCTTGGCTGGGGATGAGTCCAGGCTG 892

      306 SerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
          |||||||::|||           |||||||::::   :::|||
      891 ATGAGGTCCTTACGG...........AAGGGGCTGTTGGGGGAGCCA 856

Seq name: /SIDSB/pubdata/genseqp/genseqp/NA1999 :AT:AAV63557

Seq documentation block:
ID   AAV63557 standard; cDNA; 4243 BP.
XX
AC   AAV63557;
XX
DT   01-FEB-1999 (first entry)
XX
DE   Islet cell antibody antigen cDNA from clone ICA 12.3.
XX
KW   Pancreatic islet cell antibody antigen; ICA antigen; diagnosis;
KW   treatment; insulin dependent diabetes mellitus; IDDM; autoantibody;
KW   autoimmune disease; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             692..1618
FT                   /*tag = a
XX
PN   US5840846-A.
XX
PD   24-NOV-1998.
XX
PF   05-MAY-1994;    94US-0239276.
XX
PR   14-JUN-1991;    91US-0715181.
PR   17-FEB-1989;    89US-0312543.
PR   04-DEC-1989;    89US-0441704.
PR   08-JUN-1992;    92US-0872646.
PR   05-MAY-1994;    94US-0239276.
XX
PA   (FARB ) BAYER CORP.
XX
PT   Rabin DU;
XX
DR   WPI; 1999-044118/03.
DR   P-PSDB; AAW80484.
XX
PT   Polypeptide antigens from pancreatic islet cells - useful for the
PT   detection and treatment of insulin-dependent diabetes mellitus
XX
PS   Example 1; Columns 35-42; 47pp; English.
XX
CC   The present sequence encodes a pancreatic islet cell antibody
CC   (ICA) antigen. The proteins are used for the diagnosis and treatment
CC   of insulin dependent (Type I) diabetes mellitus (IDDM). They are
CC   recognised by autoantibodies of the disease as IDDM is now thought
CC   to be an autoimmune disease. They were identified from probing of
CC   IDDM serum with human cDNA libraries expressing protein fragments.
XX
SQ   Sequence 4243 BP; 664 A; 1007 C; 854 G; 719 T; 0 other;


Alignment scores:
            Quality:  101.50              Length:     134
              Ratio:  1.346                 Gaps:       8
 Percent Similarity:  56.716   Percent Identity: 27.612

Alignment block:
 US-09-126-945B-2 x AAV63557/rev    ..

 Align seq 1/1  to reverse of: AAV63557  from: 1  to: 4243

      209 TrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIle.. 224
          |||   :::|||:::       ::::::   |||||||||||||||
     1241 TGGGCCGCGCCTTGTACTTGTAGTCAGGATACTTCTCCAGCTGCTGCCGG 1192

      225 ....................HisTyrCysAlaSerThrSerGlaGluSerI 245
                               |||      |||:::::          ||||
     1191 CTCAGCCGCGCCCTGTTCCTCATAGTAGGGCTGCTTCTCCTGGTTGGTCAT 1142

      246 rpThrAspSerGluValAspSerSerCys.SerGlyGlnProIleHisLe 251
          |||||:::||||||:::::::::|||||| ::::::      :::|||::
     1141 GGACTTCCAGCGAGATCCAAGGATCTTGCTGATGCTGGAGTTGTGCATGT 1092

      251 rTrpGlnPheLeuLysGlu.....LeuLeuLeuLysProHisSerI 265
          :|||:::   |||:::::::         |||::|||   ||||||
     1091 CTGGGAAGGCTTGCAGGATCTTCCTCCGCTCATCCTTGGCCCACACCATG 1042

      265 yrGly.....Arg.PheIleArgTrp...LeuAsnLysGluLysGlyIl 278
          |||          ::::::   |||   |||   :::::::::|||
     1041 AAGGCGTTCATGGGCCTCTTGATGTGGCTGCTGTTCGGGACTCGGGGAA 992

      278 rPheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgL 295
          :::::::      :::   :::|||::::::         :::
      991 GTGGCGGGAGCCATCCATGTCGCAGCTCAGCATGGCTTCCTCCAGTGGGT 942

      295 ysAsnArgPro.....AlaMetAsnTyrAspLysLeu 305
          :::::|||          |||   :::   :::::|||
      941 GCACACAGCCGTCCTCCAGCCGCTCCTTGGCTGGGGATGAGTCCAGGCTG 892

      306 SerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
          |||||||::|||           |||||||::::   :::|||
      891 ATGAGGTCCTTACGG...........AAGGGGCTGTTGGGGGAGCCA 856

Seq name: /SIDSB/pubdata/genseqp/genseqp/NA2000 :AT:AAZ46456

Seq documentation block:
ID   AAZ46456 standard; cDNA; 4243 BP.
XX
AC   AAZ46456;
XX
DT   13-MAR-2000 (first entry)
XX
DE   Pancreatic islet cell antigen (ICA) encoding cDNA (clone ICA 12.3).
XX
KW   Pancreatic; islet cell antigen; insulin dependent diabetes mellitus;
KW   IDDM; ICA; immunoglobulin; T cell; B cell; islet cell destruction;
KW   human; ss.
XX
OS   Homo sapiens.
XX
PN   US6001804-A.
XX
PD   14-DEC-1999.
XX
PF   06-JUN-1995;    95US-0468577.
XX
PR   05-MAY-1994;    94US-0239276.
```

```
PR   14-JUN-1991;   91US-0715181.
PR   08-JUN-1992;   92US-0872646.
PR   17-FEB-1989;   89US-0312543.
PR   04-DEC-1989;   89US-0441703.
XX
PA   (FARB ) BAYER AG;
XX
PI   Rabin DU;
XX
DR   WPI; 2003-071667/06.
DR   P-PSDB; AAY49926.
XX
PT   Nucleic acids encoding one or more pancreatic islet cell antigens,
PT   useful in the diagnosis and treatment of insulin-dependent (type I)
PT   diabetes mellitus -
XX
PS   Claim 1; Columns 35-42; 62pp; English.
XX
CC   The invention provides cloned nucleic acids that code for one or more
CC   pancreatic islet cell antigens (ICA) or their fragments which bind
CC   selectively with ICA antibodies. The cloned nucleic acids
CC   (AAZ46451-460) are characterized by cDNA inserts in deposited recombinant
CC   bacteriophages ATCC 40550, ATCC 40553, ATCC 40554, ATCC 40561, ATCC
CC   40562, ATCC 40703, ATCC 40704, ATCC 40705, ATCC 40706, and ATCC 75000
CC   respectively. The native ICA proteins or their fragments are useful in
CC   the diagnosis of insulin-dependent (Type I) diabetes mellitus (IDDM) and
CC   in detecting or blocking human immunoglobulin, T-cells or B-cells
CC   involved in IDDM. The molecular cloning of ICA proteins also allows for
CC   the preparation of large and reproducible amounts of material for use in
CC   research, diagnosis and treatment of IDDM, as well as the opportunity to
CC   the study biological mechanisms involved in islet cell destruction and
CC   the appearance of ICA.
XX
SQ   Sequence 3243 BP; 664 A; 1007 C; 853 G; 719 T; 0 other;


alignment_scores:
            Quality:  101.50              Length:     134
              Ratio:  1.336                 Gaps:       8
 Percent Similarity:  56.716   Percent Identity: 27.612

alignment_block:
 US-09-126-945B-2 x AAZ46456/rev   ..

 Align seq 1/2  to reverse of  AAZ46456  in  m  1  *    3213

     209 TrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIle.. 224
         |||   ::|||:::       ::::::   ||||||||||||||||
    1242 TGGGCCGCGGCTTGTACTTGTAGTCAGGATACTTCTCCAGGTGCTGCCGG 1192

     225 .....HisTyrCysAlaSerThrSerGluGluSer 245
                    |||      |||:::::         ||||
    1191 CTCAGCCGCGCCTGTTCCTCATAGTAGGGCTGCTTCTCCTGGTTGGTCAT 1142

     246 rpThrAspSerGluValAspSerSerCys SerGlyGlnProIleHisIle 261
         |||:::|||||||:::::::::|||||| ::::::      :::|||::
    1141 GGACTTCCAACGAGATCCAAGGATCTTGCTGATGCTGGAGTTGTGCATCT 1092

     251 uTrpGlnPheLeuLysGlu.........LeuLeuLeuLysProHisSerT 265
         :|||:::   |||::::::          |||::::||   ||||||
    1091 CTGGGAAGGCTTGCAGGATCTTCCTCCGCTCATCCTTGGCCCACACCATG 1042

     266 yrGly......Arg.PheIleArgTrp...LeuAsnLysGluLysGlyIl 278
           |||       ::::::   |||   |||   ::::::::::|||
    1041 AAGGCGTTCATGGGCCTCTTGATGTGGCTGCTGTTTCGGGACTCGGGGAA 992

     278 ePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgL 295
         ::::::      :::   :::|||::::::         :::
     991 GTGGCGGGAGCCATCCATGTCGCAGCTCAGCATGGCTTCCTCCAGTGGGT 942

     295 ysAsnArgPro.................AlaMetAsnTyrAspLysLeu 305
         ::::::|||                 |||  :::  ::::::|||
     941 GCACACAGCCGTCCTCCAGCCGCTCCTTGGCTGGGGATGAGTCCAGGCTG 892

     306 SerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
         ||||||::|||           ||||||:::::   :::|||
     891 ATGAGGTCCTTACGG.............AAGGGGGTGTTGGGGGAGCCA 856

seq_data  [ZIPSS:Sequence_data_seq_number_NAZ46456_AAZ29482

seq_documentation_block:
ID   AAZ29482 standard; DNA; 1192 BP.
XX
AC   AAZ29482;
XX
DT   14-MAR-2000  (first entry)
XX
DE   DNA encoding human cell cycle regulation protein-3.
XX
KW   CECRP-3; cell cycle regulation protein-3; tissue expression; cancer;
KW   inflammation; pBluescript vector; gene therapy; disease diagnosis;
KW   disease monitoring, chromosomal mapping, cell proliferation;
KW   cell proliferative disorder; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             140..850
FT                   /*tag= a
FT                   /product= "Cell cycle regulation protein-3"
XX
PN   WO9964594-A2.
XX
PD   16-DEC-1999.
XX
PF   08-JUN-1999;   99WO-US12906.
XX
PR   08-JUN-1998;   98US-0088695.
XX
PA   (INCY ) INCYTE PHARM INC.
XX
PI   Bandman O,  Lal P,  Tang YT,  Corley NC,  Guegler KJ,  Baughn MR;
PI   Patterson C;
XX
DR   WPI; 2000-105887/09.
DR   P-PSDB; AAY44362.
XX
PT   New regulatory protein, useful in diagnosis, treatment and prevention of
PT   cell proliferative and immune system diseases -
XX
PS   Claim 7; Page 75; 88pp; English.
XX
CC   The present sequence encodes cell cycle regulation protein 3 (CECRP 3).
CC   Tissue expression is seen in reproductive (0.302), nervous (0.164) and
CC   cardiovascular (0.186) tissues. The cDNA is cloned into pBluescript
CC   vector. Diseases associated with tissue expression of CECRP-3 are cancer,
CC   inflammation and other diseases. This is used in the recombinant
CC   production of CECRP-3, in gene therapy (including expression of antisense
CC   sequences, triplex-forming molecules or ribozymes), and as source of
CC   probes and primers for diagnosis and monitoring of disease. It is also
CC   used for chromosomal mapping and isolation of related sequences. CECRPs
CC   are activators of cell proliferation or inhibitors of cellular processes
CC   that modulate proliferation. They are used to treat cell proliferative
CC   disorders.
XX
SQ   Sequence 1192 BP; 237 A; 350 C; 375 G; 230 T; 0 other;


alignment_scores:
            Quality:  101.00              Length:     258
              Ratio:  0.878                 Gaps:      16
 Percent Similarity:  44.574   Percent Identity: 24.041

alignment_block:
 US-09-126-945B-2 x AAZ29482   ..
```

```
Align seq 1/1  to: AAZ29482  from: 1  to: 1192

      15 HisLeuLeuLeuProProAspThrValSerArgThrGlyLeuGluLysAl  31
         ||||||:::|||     :::|||:::   |||       :::   ||
     335 CACCTGGTGCTGGTCCTGAACACTCTGCGGCGCATCCAGGCGTCCATGGC  384

      41 aAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProP  48
         |   |||:::|||::                          ||||||||||
     385 ACCCGGCGCTGCCCTG..............CCACCTGTGCCTAGCCCAC  419

      48 roAlaThrPro..............................GluGlnGly  54
         |||||:::|||                          :::   :::
     420 CTGCAGCCCCCAGTGTGGCTGACAACTTACTGGCAAGCTCGGACGCTGCC  469

      55 LeuSerAlaPheTyrLeuSerTyrPhe...AspMetLeuTyrProGluAs  70
         |||||||||          |||  :::   |||:::   :::   |||
     470 CTTTCAGCCTCCATGGCCAGCCTCCTGGAGGACCTCAGCCACATGAGGG  519

      70 pSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGlnGluProP  87
           |||         :::|||||||          :::   |||   ||||
     520 CCTGAGT.........CAGGCTCCCCAACCCTTGGCAGACGAGGGGCCAC  560

      87 roGluGluProGluGlnCysProValIleAspSerGlnAlaProAla...  102
         ||   :::                  |||   :::   |||||||:::
     561 CAGGCCGTAGC..............ATCGGGGGAGCAGCGCCCAGCCTG  595

     103 GlySerLeuAspLeuVal...ProGlyGlyLeuThrLeuGluGluHisSe  119
         |||:::||||||||||:::   |||:::         |||   :::   ::
     596 GGTGCCTTGGACCTGCTGGGCCCAGCCACTGGCTGTCTACTGGACGATGG  645

     118 rLeuGluGlnValGluSerMetValValGlyGluValLeuLysAspIleG  135
         :||||||                          ::::::::::||||||
     646 GCTTGAG.......................GGCCTGTTTGAGGATATTG  671

     135 luThrAla.CysLysLeuLeuAsnIleThrAlaAspProMet........  148
         ::|||||: |||      :::|||:::      :::|||:::
     672 ACACCTCTATGTATGACAATGAACTTTGGGCACCAGCCTCTGAGGCCTC  721

     149 ..............................................AspTrpS  153
                                                     :::|||:
     722 AAACCAGGCTCTGAGAtGGcCGGGCAAGGAGGAAGCTCCCGACGTGGA  771

     151 er...ProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyr  166
         ::   |||:::   :::   :::|||:::   |||      ||||||||:
     772 CGAGcCCAATGGACTACCTCATGGATGTGCTGGTGGGCACACAGGCAC  821

     167 ArgLeuProProMetGlyLysAlaPheGlnGlu.................  177
                   |||||||||   :::
     822 TGGAGCGACCGCCGGGGCCAGGGCGCTGAGCCCTCGTGCTGGAATGGTTG  871

     178 ......................LeuAlaGlyLysGluLeuCysAlaMetS  187
                              |||||||||      :::      :
     872 TCTGGTATCTGAACTGACCCTGCTGGCTGGACCAACTGTCCTCGAAAAGA  921

     187 erGlu...GlnGlnPheArgGlnArgSerProLeuGlyGlyAspValLeu  202
         :::::      ::::::   ||||||:::   ||||||
     922 CACAGCTGGCTTCCCTAGTACAGAGAACAGGGCTTGGG............  959

     203 HisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgLI  219
                    :::|||                          ::::::|||
     960 ............CCACTTTGG.................AGAGACAGAAT  979

     219 hrSerProGlyAlaIleHis  225
         ::||||||||||   :::|||
     980 CTAGTCCTGGGCAACTTCAC  999

Seq name: /cgi30/1_dir/1/seq -1 /tmp -p /NA/08.DAT:AAV18886

seq documentation block:
ID   AAV18886 standard; cDNA; 8575 BP.
XX
AC   AAV18886;
XX
DT   17 AUG 1998  (first entry)
XX
DE   Homo sapiens von Willebrand factor cDNA.
XX
KW   PCR primer; phage vector; bacteriophage lambda; gene therapy;
KW   antisense RNA; ribozyme; delivery; vaccine; antigen; treatment;
KW   prevention; in vivo; in vitro; AIDS; cancer; cystic fibrosis;
KW   acquired immune deficiency syndrome; Gaucher's disease; haemophilia;
KW   post translational modification; diagnostic reagents; production;
KW   von Willebrand factor; ss.
XX
OS   Homo sapiens.
XX
PN   WO9805344 A1.
XX
PD   12 FEB-1998.
XX
PF   03 JUL-1997;   97WO-US12928.
XX
PR   11 MAR-1997;   97US-0814859.
PR   05 AUG-1996;   96US-0693865.
XX
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX
PI   Dubin DB,  Kupper IS,  Sarkar SN;
XX
DR   WPI; 1998-145445/13.
XX
PT   Improved gene therapy with phage vectors unable to release genetic
PT   material into cells - comprises surface ligand binding to
PT   cell surface receptor, and transcribing nucleic acid in cell, useful
PT   for, e.g. treating AIDS
XX
PS   Example 8; Pages 77-80; 110pp; English.
XX
CC   The sequence is that of human von Willebrand factor cDNA which
CC   was used in construction of a chimeric bacteriophage lambda vector.
XX
SQ   Sequence 8575 BP; 1759 A; 2471 C; 2555 G; 1790 T; 0 other;


alignment scores:
            Quality:  101.00              Length:    287
              Ratio:  0.842                 Gaps:     11
 Percent Similarity:  41.812   Percent Identity:  22.360

alignment block:
 US-09-126-945B-2 x AAV18886  ..

 Align seq 1/1  to: AAV18886  from: 1  to: 8575

       2 GlySerAlaSerProGlyLeuSerSerValSerPro......SerHisLe  16
         |||   ||||||||      :::|||:::   |||      |||
    4016 GGCTGCGCATCCCAGAAGTGGGTCGGCGTGGCCGTGGTGGAGTACCAC 4065

      16 uLeuLeuPro...ProAspThrValSerArgThrGlyLeuGluLysAlaA  32
           |||   |||   :::   ||||||||||||   ::::::   :
    4066 GACGGCTCCCACGCCTACATCGGGCTCAAGGACCGGAAGCGACCGTCAGA 4115

      32 laAlaGlyAlaVal............GlyLeuGluArgArgAspTrpSer  44
         :::::||||||:::          :::::::::          |||
    4116 GCTGCGGCGCATTCCAGCCAGGTGAACTATGCGGGCAGCCACGTGC... 4162

      45 ProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSe  61
           |||||||||
    4163 ...CCTCCACCAGCGAGG.............................. 4177

      61 rTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaP  78
                          :::   ::::::   ::::::||:::
    4178 .....................TCTTGAAATACACACTGTTCCAAATCTT 4205

      78 roGlyAlaSerSerArgGluGluProProGluGluProGluGlnCysPro  94
           :::   |||:::      :::|||||||   :::|||      |||
```

```
     4206 CAGCAAGATCGACCGCCCTGAAGCCTCCCGCATCGCCC.....TGCTCC 4249

       95 ValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGl 111
                  :::::::::::|||   |||                 ||||||
     4250 TGATGGCCAGCCAGGAGCCCCAACGGA.............TGTCCCGGAAC 4287

      111 yLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValG 128
          ||||::      :::::::::   ::::::   :::|||:::
     4288 TTTGTCCGCTACGTCCAGGGCCTGAAGAAGAAGAAGGTCATTGTGATCCC 4337

      128 lyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThr 144
            ::::||      :::   ||||::   :::      :::
     4338 GGTGGGCATTGGGCCCCATGCCAACCTCAAGCAGATCCCCCTCATCGACA 4387

      145 AlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTr 161
          :::   ||::::        ||||||            ||::::   ||
     4388 AGCAGGCCCCTGAGAACAAGGCCTTCGTGCTGAGCAGTGTGGATGAGCTG 4437

      161 pThrGluHisGlnTyrArg.........................LeuP 169
          |:::::        |||                          |||
     4438 GAGCAGCAAAGGAGCAGATCGTTAGCTACCTCTGTGACCTTGCCCCTGA 4487

      169 roProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAla 185
            ||:::                        ||::::|||   |||
     4488 AGCCCCTCCTCCTACTCTGCCCCCCACATGGCACAAGTCACTGTGGGCC 4537

      186 MetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValIe 202
          ::::::      ||||||      ::::||   |||        ::
     4538 CGGGGTCTTGGGGGTTTCGACCCTGGGGCCCAAGAGGAACTCCATGTTT 4587

      202 uHisAlaHisLeuAspIleTrpLysSer..................... 211
          :       :::   ||||||:::
     4588 CTGGATGTGGCCGTTCGTCCTGCAAGGATCGGACAAAATTGGTGAAGCCA 4637

      212 .....................................AlaAlaTrpMet 215
                                             ::::::||||||
     4638 CTTCAACAGGAGCAAGGAGTTCATGGAGGAGGTGATTCAGCGGATGGATG 4687

      216 LysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThr..... 230
            ||||||:::      :::        ||:::|||||
     4688 TGGGCCAGGACAGCA......TCCACGTCACGGTGCTGCAGTACTCCTAC 4731

      231 .............SerGluGluSerTrpThrAspSerGluValAspSerS 243
                      |||     |||      |||       :::|
     4732 ATGGTGACCGTGGAGTACCCCTTCAGCGAGGCACAGTCCAAAGGGGACAT 4781

      243 erCysSerGly 246
          ||| |||||||
     4782 CCTGCAGCGGG 4792

seq name: US126945b2gbrseq@pctoseq@petosc.pc_NA1006_0A1:AAN60404

seq_documentation_block:
ID   AAN60404 standard; cDNA; 8585 BP.
XX
AC   AAN60404;
XX
DT   25-JUN-1991  (first entry)
XX
DE   Sequence encoding human von Willebrand Factor (VWF) precursor.
XX
KW   Chronic renal failure; therapy; factor VIII C; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/qualifiers
FT   CDS             107..8548
FT                   /*tag= a
XX
PN   WO8606096-A.
XX
PD   23-OCT-1986.
XX
PF   10-APR-1986;    86WO-US00760.
XX
PR   11-APR-1985;    85US-0722108.
XX
PA   (CHIL-) CHILDRENS MED CENT.
PA   (GINS/) GINSBURG D.
XX
PI   Ginsburg D, Orkin SH, Kaufman RJ;
XX
DR   WPI; 1986-291663/44.
DR   P-PSDB; AAP60462.
XX
PT   Pure Von Willebrand Factor - produced using an expression vector
PT   including a DNA sequence encoding functional VWF protein
XX
PS   Claim 2; Table 2, Pages 18-36A; 54pp; English.
XX
CC   cDNA clones pVWH33, pVWH5 and PVWE6 which span 9 kb pairs of DNA and
CC   encompass the entire protein coding region of VWF, were selected to
CC   construct full length cDNA (AAN60404). The pure VWF produced is useful
CC   in the treatment of von Willebrand's disease (VWD) and the patients
CC   with chronic renal failure whose abnormal bleeding times are
CC   corrected by crude cryoprecipitate. Pure VWF can also be used to
CC   carry, stabilise and improve the therapeutic efficacy of factor
CC   VIII:C.
XX
SQ   Sequence 8585 BP; 1762 A; 2472 C; 2555 G; 1796 T; 0 other;


alignment_scores:
            Quality:  101.00              Length:    287
              Ratio:  0.842                Gaps:     11
 Percent Similarity:  41.812   Percent Identity: 22.300

alignment_block:
 US-09-126-945B-2 x AAN60404   ..

 Align seq 1/1  to: AAN60404  from: 1  to: 8585

       2 GlySerAlaSerProGlyLeuSerSerValSerPro.....SerHisLe 16
         |||   |||||||||        :::|||:::   |||      |||
    4023 GGCTGGGCATCTCCCAGAAGTGGGTCCGCGTGGCCGTGGTGGAGTACCAC 4072

      16 uLeuLeuPro...ProAspThrValSerArgThrGlyLeuGluLysAlaA 32
           |||   |||   :::   |||||||||||||   ::::::   :
    4073 CACGGCTCCCACGCCTACATCGGGCTCAAGGACCGGAAGCGACCGTCACA 4122

      32 laAlaGlyAlaVal.............GlyLeuGluArgArgAspTrpSer 44
         ::::|||||:::             :::::::::          |||
    4123 GCTGCGGCGCATTGCCAGCCAGGTGAAGTATGCGGGCAGCCAGGTGG... 4169

      45 ProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSe 61
           |||||||||
    4170 ...CCTCCACCAGCGAGG.............................. 4184

      61 rTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaP 78
                              :::   ::::::   ::::::|||:::
    4185 ......................TCTTGAAATACACACTGTTCCAAATCTT 4212

      78 roGlyAlaSerSerArgGluGluProProGluGluProGluGlnCysPro 94
         :::   |||:::      :::|||||   :::|||      |||
    4213 CAGCAAGATCGACCGCCCTGAAGCCTCCCGCATCGCCC.....TGCTCC 4256

      95 ValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGl 111
               :::::::::::|||   |||                 ||||||
    4257 TGATGGCCAGCCAGGAGCCCCAACGGA.............TGTCCCGGAAC 4294

     111 yLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValG 128
         ||||::      :::::::::   ::::::   :::|||:::
    4295 TTTGTCCGCTACGTCCAGGGCCTGAAGAAGAAGAAGGTCATTGTGATCCC 4344

     128 lyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThr 144
           ::::||      :::   ||||::   :::      :::
```

```
  4345 GCTGGGCATTGGGCCCCATGCCAACCTCAAGCAGATCCGCCTCATCGACA 4394

   145 AlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTr 161
       :::   ||:::        ||||||           ||:::   ||
  4395 AGCAGGCCCCTGACAACAAGGCCTTCGTGCTGAGCACTGTGGATGACCTG 4444

   161 pThrGluHisGlnTyrArg.....LeuP 169
       ::::::        |||                |||
  4445 GAGCAGCAAAGGACGAGATCGTTACGTACCTGTGACCTTGCCCCTGA 4494

   169 roProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAla 185
         ||:::                 |||:::|||   |||
  4495 AGCCCCTCCTCCTACTCTGCCCCCCCACATGGCACAAGTCACTGTGGCCC 4544

   186 MetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLe 202
       ::::::      ||||||     :::|||   |||          ::
  4545 CGGGGCTCTTGGGGGTTTCGACCCTGGGGCCCAAGAGGAACTCCATGGTT 4594

   202 uHisAlaHisLeuAspIleTrpLysSer..... 211
       :          :::   ||||||||:
  4595 CTGGATGTGGCGTTCGTCCTGGAAGGATCGGACAAAATTGGTGAAGCCGA 4644

   212 .....AlaAlaTrpMet 215
                :::::|||||
  4645 CTTCAACAGGAGCAAGGAGTTCATGGAGGAGGTGATTCAGCGGATGGATG 4694

   216 LysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThr..... 230
       |||||||:::      :::      |||:::||||||
  4695 TGGGCCAGGACAGCA.....TCCACGTCACGGTGCTGCAGTACTCCTAC 4738

   231 .............SerGluGluSerTrpThrAspSerGluValAspSerS 243
                   |||     |||          |||          :::|
  4739 ATGGTGACCGTGGAGTACCCCTTCAGCGAGGCACAGTCCAAAGGGACAT 4788

   243 erCysSerGly 246
       ||||||||||
  4789 CCTGCAGCGGG 4799

Seq name: gi|[illegible]|gb|[illegible] Homo sapiens cDNA, 4799 bp

seq documentation_block:
ID   AAF58162 standard; DNA; 1246 BP.
XX
AC   AAF58162;
XX
DT   24-APR-2001 (first entry)
XX
DE   Human neuron progenitor cell clone DNA.
XX
KW   Neuron; progenitor cell; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200107607-A2.
XX
PD   01-FEB-2001.
XX
PF   21-JUL-2000; 2000WO-JP04895.
XX
PR   23-JUL-1999;   99JP-0209817.
PR   18-OCT-1999;   99US-0159628.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Kawai Y;
XX
DR   WPI; 2001-182791/18.
XX
PT   New human polynucleotides, particularly DNAs, isolated from a cDNA
PT   library derived from progenitor cells, useful in gene therapy, as well
PT   as in producing proteins useful as diagnostic markers in drug
PT   development
XX
PS   Claim 1; Page 28-31; 54pp; English.
XX
CC   The present invention relates to human proteins isolated from clones
CC   from neuron progenitor cells. The proteins and the DNA encoding them may
CC   be used in the preparation of treatments for diseases associated with
CC   the proteins.
XX
SQ   Sequence 1246 BP; 235 A; 346 C; 486 G; 279 T; 0 other;


alignment scores:
            Quality:  100.50              Length:     376
              Ratio:  0.602                 Gaps:      21
 Percent Similarity:  44.415   Percent Identity: 22.074

alignment block:
 US-09-126-945b-2 x AAF58162   ..

 Align seq 1/1  to: AAF58162  from: 1  to: 1246

        7 GlyLeuSerSerValSerProSerHisLeuLeuLeuProProAspThrVa 23
          |||   ::::::   :::   |||:::   :::   |||   ::::::::
      176 GGCGCCACGGAGAGGAACCGCTCTAGGCACGTAAGGCCTCCGTGAGGTGC 225

       24 lSerArgThr.....GlyLeuGluL 40
          :|||||||::                |||   :::
      226 GTGGCGGCGGCGGAGCACTCTGGGACTTGTAGTTCTGGAGATGGAGCGAGCT 275

       40 ysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSer 46
          |:|||:   ||||||   |||   ::::|||
      276 GTGCCGCTCGCCGGTGCCTCTGGGTCAGACAGA..... 307

       47 ProProAlaThrProGluGluGlyLeuSerAlaPheTyrLeuSerTyrPh 63
                                 |||:::   ::::::      :::   ::
      308 .....AGTCCTCCAGGCCTTGCAGCAGCTCCATA 336

       64 eAspMetLeuTyrProGluAsp...SerSerTrpAlaAlaLys.AlaPro 78
          :|||   |||   ||||||      :::   |||   :::::: |||
      337 TGACCATCTTCTCCAGAGCGTCTCACCATGTGGGAAGTTTCTGGCGGCT 386

       79 Gly.....AlaSerSerArgGl 84
          |||                 ::|||:::   ::
      387 GGCAACAAATTACGAGAGATGCCATCTTCAGCTTGTCCTCTGCTTTGAG 436

       84 uGluProProGluGluProGluGlnCysProValIleAspSerGluAlaP 101
          ::::      :::|||   ::::::   |||||||::       ||||||
      437 CTCAGAAGCCAAAGAGGAAAGTAAGAAGCCGGTGGTGACTTTCCAAGCCC 486

      101 roAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHis 117
            |||           |||   :::   ::::::   ::::::|||
      487 ATGATGGGCCC.....GTCTATAGCATGGTTTCCACCGATCGACAT 527

      118 SerLeuGlu.....GlnValGluSerMetValValGlyGluVa 130
          |||:::          :::   ::::::   :::   :::||:::
      528 CTGCTTAGTGCTGGGGATGGGGAGGAGAAGGCCTGGCTTGGGCGGAGAT 577

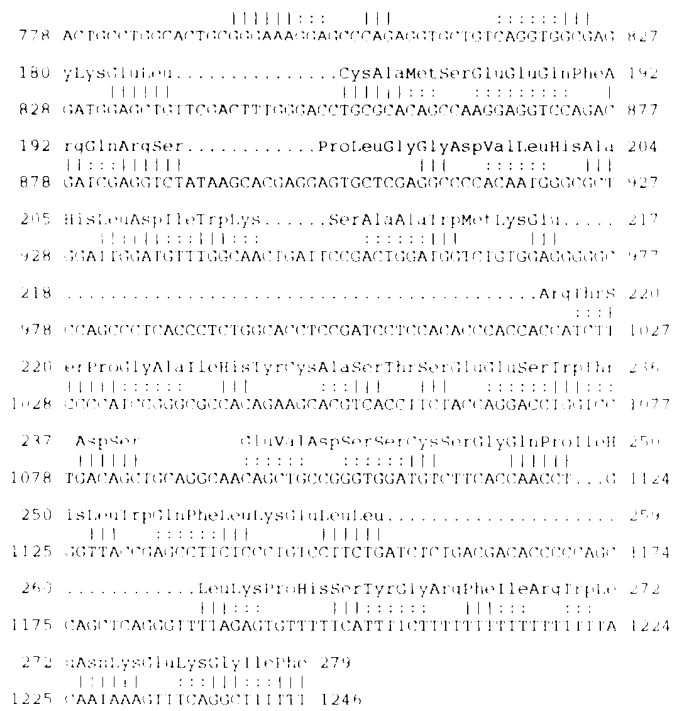
      130 lLeuLysAsp.IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAsp 146
          :||||||          ::|||   ::::::::::   :::
      578 GCTCAAGAAGGGCTGTAAGGAGCTGTGGCGTCGTCAGCCTCCATACAGGA 627

      147 Pro.....MetAspTrpSerProSerAs 154
          |||                 |||||||||   ::
      628 CCAGCCTGGAAGTGCCTGAGATCAACGCTTTGCTGCTGGTCCCCAAGGAG 677

      154 nVal...GlnLysTrpLeu.....LeuTrp.... 161
          ::::   :::::||||||                |||||
      678 AATTCCCTCATCTGGCTGGGGGAGACTGTCAGTTGCACACTATGGACCT 727

      162 .....ThrGluHisGlnTyr 166
                                 |||:::
      728 TGAAACTGGGACTTTCACGAGGGTCCTCCGGGGCCACACAGACTACATCC 777

      167 ArgLeuProProMetGlyLysAlaPheGln.....GluLeuAlaGl 180
```



```
BLOSUM Matrix: BLOSUM62 Issued Patents NA *  Gap_Penalty
Date: Sep 18, 2001  7:10 AM
About: Results were produced by the GenCore software version 4.5
       Copyright (c) 1993-2001 Compugen Ltd.
Command line parameters:
  MODEL=frame+_p2 model  PFP x3
  Query file (SEQ  specified sequence 945b-2.rni  17 pairs  11 file  Gap 8  Ext 2
  DB Issued Patents NA  QFMT fastap  PREFIX rni  SAP 8 32 4 0
  -GAPEXT=4.000 -MINMATCH 0.100  LOOPCT 0.000  XSAPPXT 0.000
  -GAPOP 4.500 -SGAPPXT 0.050  XSAP=4 15 0.5  XSAPEXT 0.5
  -EGAP 8.5  EGAPEXT 7.500  YSAP 8 15 0.5  YSAPEXT 0.5
  -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX blosum62
  -TRANS=human40.cdi -LIST=150  LOCATION 200  PRF 8.0=EE 1.0
  -THR MAX 1.0  DB_MIN 0  ALIGN 120  MODE 1 *AL  -SUBMIT pir
  NORM=opt  MODE=FN  MAXDB 2000000000
  -USER=US09126945 @GN1_1_60  NCPUS 6 -ICPU=3 -LONG=0 -N= Y=50
  -WAIT -THREADS 1

Search Information block:
Query: US-09-126-945B-2
Query length: 825
Database: Issued Patents NA *
Database sequences: 871554
Database length: 94655562
Search time (sec): 69.070000

score_list:
Sequence          Std Orig    %Sim to    %Sim to Len  Length  Score  %
```

/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-441-677-1 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-454-543-1 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-414-967-1 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-458-076A-6 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-457-812-6 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-457-646A-6 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-458-076A-6 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-764-248A-4 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-457-395A-6 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-729-214-6 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-028-944-6 * 88.50 90.06 2.1e-03 28958 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-862-134-13 * 88.50 128.42 15.87 910 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-657-866-10 * 89.00 127.44 17.54 1095 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-720-484A-8 * 88.00 115.11 85.29 2972 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-824A-8 * 88.00 115.11 85.29 2972 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-440-011-1 * 88.00 111.19 140.99 4195 !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-901-710-1 * 88.00 111.19 140.99 4195 !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-795-430-1 * 88.00 110.60 151.95 4416 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-222-636-31 * 88.00 110.58 152.41 4425 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-04228-31 * 88.00 110.58 152.41 4425
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-642-111-7 * 88.00 109.62 171.44 4795 !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-901-710-8 * 88.00 109.62 171.44 4795 !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-896-449A-1 * 89.50 102.29 415.54 8802 !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-132-652-1 * 88.00 102.76 415.54 8802 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US96-12648-47 * 88.00 102.76 415.54 8802 !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-188-940-17 87.50 138.62 4.18 349 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-07-841-591A-1 * 92.00 127.12 18.28 960 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US94-02945-1 * 92.00 127.12 18.28 960
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-086-601-2 * 87.50 124.58 25.41 1200 !
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-908-548-32 * 87.50 124.58 25.41 1200 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-468-411-32 * 87.50 124.58 25.41 1200 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US96-12648-32-8 * 87.50 124.58 25.41 1200 ! Pate
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-451-777A-6 * 87.50 123.25 30.03 1349 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-451-778A-5 * 87.50 123.25 30.03 1349 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-908-548-6 * 87.50 123.25 30.03 1349 !
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-06743-6 * 87.50 123.25 30.03 1349

Seq name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-009-913-8

Seq documentation block:

; Sequence 8, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: AxyS Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,913
; FILING DATE: 21-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ 4P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650 327 3241
; TELEFAX: 650 327 3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-009-913-8

Alignment_scores:
Quality: 300.50 Length: 262
Ratio: 2.013 Gaps: 6
Percent Similarity: 56.870 Percent Identity: 30.153

Alignment_block:
US-09-126-945B-2 x US-09-009-913-8 ..

Align Seq 1/1 to: US-09-009-913-8 from: 1 to: 2280

1o2 AraGlySerLeuAspLeuValProArgGlyLeuThrLeu..... 114
||| :|||||||||||||||:: :::::|||||||:::
32 GCCTCCTCTTTGGACCTAGCCACCGCTGCCCTCACGGTAATGTTGGACTC 81

.15 .GlnGlnAlaSerGlnGlnGlnValGluSerMetValValArgGluVal1L 131
|||||| |:::::: :::: :
82 CGTGACACACAGCACCTTCCTGCCTAATGCATCCTTCTGCGATCCCCTGA 131

131 cuGys.....Asp 133
::::: :::
132 TGTCGTGCAGTGATCTGTCAGCAATGAAGAGTACTACCCTGCCTTTGAG 181

133 IleGlnThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIr 150
:::|||||||| :::::: ||| ||
182 ATAGAGACACCGTGACTCATACTGGACATCAGTCCACCCTGAATACTG 231

150 pSerProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrA 167
|::: :::||| :::|||||| ::: |||||||
232 GACTAAGCGCCATGTGTGGGAGTGGCTCCACGTTCTGCTGCGACCAGTACA 281

167 rqLeuProProMetGlyLysAlaPhe.....GlnGluLeuAlaGlyLys 181
::|||| :::||| ||||||||||
282 AGTTGGACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGGCCTG 331

182 GluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGl 198
:::|||||||:::|||:::::|||:::||| ::: ::: |||
332 CAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAGCTGGCTCTG 381

198 yGlyAspValLeuHisAlaHisLeu..... 206
|||::: |||::: |||
382 CGGCGAGTACCTGTACTTCATCCTCCAGAACATCCGCACACAAGGTTACT 431

207 .....AspIleTrpLysSerAlaAlaTrpMetLysGluArgThr 219
||| :::||| ||| :::|||::: :::
432 CCTTTTTAAATGACGCTGAAGAAAGCAAGGCCACCATCAAAGACTATGCT 481

220 SerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpTh 236
::: ::: ||| :::|||||| :::
482 GATTCCAAC.....TGCTTGAAAACAAGTGGCATCAAAAGTCA 519

236 rAspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpG 253
||| ::: ||| ||| ||| |||||||||:
520 AGACTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGG 569

254 lnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 269
::|||:::||||||||||:::||||::::: ::::::
570 AATTTGTACGACACTGCTTCTATCTCCTGAACAAAACTGTGGCATTCTG 619

270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 286
:::||| :::::::|||::|||||||||:::::: |||
620 GAATGGGAAGATAGGGAACAAGGAATTTTTCGGGTGGTTAAATCGGAAGC 669

```
     286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 303
         :::|||::::::::|||||   ||||||:::::::      |||::::|||:
     670 CCTGGCAAAGATGTGGGGACAAAGGAAGAAAAATGACAGAATGACGTATG 719

     303 spLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArg 319
         :::|||||||||||||:::::::|||   |||||||||   ||||||:::::
     720 AAAAGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGGAATTITGGAG 769

     320 LysProAspIleSerGlnArgLeuValTyrGlnPhe 331
         :::   |||      :::|||||||||||||:::|||
     770 CGGGTTGAC.....CGAAGGTTAGTGTACAAATTT 799

seq_name: /cygd2_6/ftsdata/1/ina/6A_COMB.seq;US-09-009-913-10

seq_documentation_block:
; Sequence 10, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;    APPLICANT:  AxyS Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  Asthma Related Genes
;    NUMBER OF SEQUENCES:  339
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION.
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  SEQ-4P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650 327-3231
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2498 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-009-913-10


 alignment_scores:
             Quality:  299.50              Length:    350
               Ratio:  1.576                Gaps:     14
 Percent Similarity:  54.286   Percent Identity: 29.429

 alignment_block:
  US-09-126-945B-2 x US-09-009-913-10   ..

  Align seq 1/1  to: US-09-009-913-10  from: 1  to: 2498

        3 SerAlaSerProGlyLeu.SerSerValSerProSerHisLeuLeuLeuP 19
          :::|||||||   ||||||     :::::::::   ||||||      :::|
       85 GCGGCCTCTGGAGGCCTGCCATGGCACCCAGAGCAGCCAT.....TTTC 129

       19 roProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAla 35

                                                  |||      ||||||
    ||||||::::::                              .....GGGGCTTTGGAAGGAGCT 157

     36 ValGly.LeuGluArgArg......AspTrpSerProSerProProAlaT 50
        |||    ||||||:::        :::   |||   ||||||||||::::
    158 TGCGGATCAGGAGAGGGAGCCTCCGCAGGGCTCTGGCTCCCCTCCAGGGG 207

     50 hrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeu 66
        ::          :::|||                              |||
    208 CCGAGGCCGCACACAAAGCCGCT.......................CTG 233

     67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 83
        :::|||:::   :::   |||   :::          |||
    234 TGGCCCAATTACACCTACTGGATAGGATTGTTGAGGGGA........... 272

     83 gGluGluProProGluGluProGluGlnCysProValIleAspSerGluA 100
                |||||||:::   |||          :::   ::::::::::|
    273 .........CCTGAGAAACTTGAGACCACAAGAACGCGTAGCGCCTCGG 312

    100 laProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGlu... 115
        ||      |||          ||||||      :::|||  ||||::
    313 CTGGCTCAGGGTGCTGAGCCCTCGTGTTGTGTTCTCTCCAGCTTCCCCG 362

    116 ......GluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
              :::|||||   :::::::::   ::::::
    363 TGCCTCAGCCACTCTTCACGTTCCATCTGTGCTCTG.............. 398

    130 lLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspP 147
                       ||||||                  ::::::  |
    399 .........TGCTGACCCGCCTGTGACTCATACTGGACATCAGTCCACC 438

    147 roMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGlu 163
        ||      |||:::      :::|||  :::|||||   :::
    439 CTGAATACTGGACTAAGCGCCATGTGTGGGAGTGGCTCCAGTTCTGCTGC 488

    164 HisGlnTyrArgLeuProProMetGlyLysAlaPhe......GlnGluLe 178
          ||||||:::|||                ::::||      :::::
    489 GACCAGTACAAGTTGGACACCAATTGCATCTCCTTCTGCAACTTCAACAT 538

    178 uAlaGlyLysGluLeuCysAlaMetSerGluGlnGluPheArgGlnArgS 195
        :::::||   :::|||||||:::|||::::::|||:::|||   :::   :
    539 CAGTGGCCTGCAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAG 588

    195 erProLeuGlyGlyAspValLeuHisAlaHisLeu............... 206
        ::   |||  |||:::  |||:::      |||
    589 CTGGCCTCTGCGGCCAGTACCTGTACTTCATCCTCCAGAACATCCGCACA 638

    207 ....................AspIleTrpLysSerAlaAlaTrpMetLy 216
                            |||      :::|||  |||   :::||
    639 CAAGGTTACTCCTTTTTTAATGACGCTGAAGAAAGCAAGGCCACCATCAA 688

    216 sGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluG 233
        |:::   ::::::  :::            |||   :::|||||
    689 AGACTATGCTGATTCCAAC....    .....TGCTTGAAAACAAGTGGCA 726

    233 luSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIle 249
          :::      |||   :::      |||      |||   |||
    727 TCAAAAGTCAAGACTGTCACAGTCATAGTACAACAAGCCTCCAAAGTTCT 776

    250 HisLeuTrpGluPheLeuLysGluLeuLeuLeuLysProHisSerTyrGl 266
        ||||||||||:::|||:::::::::::::::::::::::::::::::::
    777 CATCTATGGGAATTTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTG 826

    266 yArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluA 283
          ::::::::::|||   ::::::|||:::|||||||||:::::::
    827 TGGCATTCTGGAATGGGAAGATAGGGAACAAGGAATTTTTCGGGTGGTTA 876

    283 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 299
        |||      :::|||:::::::|||||   ||||||::::::
    877 AATCGGAAGCCCTGGCAAAGATGTGGGGACAAAGGAACAAAAATGACAGA 926

    300 MetAsnTyrArgLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
        |||:::|||:::|||||||||||:::::::|||   |||||||||   ||
```

```
      927 ATGACGTATGAAAAGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGG 976

       416 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 441
           |||:::::::::   |||       :::|||||||||||||:::|||
       977 AATTTGGAGCGGGTTGAC.....CCAAGGTTAGTGTACAAATTT 1016

Seq name: /cgn2_6/ptodata/7/ina/6A_COMB.seq Us-09-344-579-1

Seq documentation block:
; Sequence 1, Application US/09344579
; Patent No. 6054416
; GENERAL INFORMATION:
;  APPLICANT: Brenda F. Baker
;  APPLICANT: Lex M. Cowsert
;  TITLE OF INVENTION: ANTISENSE MODULATION OF ETS 2 EXPRESSION
;  FILE REFERENCE: RTS 0064
;  CURRENT APPLICATION NUMBER: US/09/344,579
;  CURRENT FILING DATE: 1999-06-25
;  NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
;  LENGTH: 2268
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (291)..(1700)
US-09-344-579-1


Alignment scores:
            Quality:  299.00              Length:     552
              Ratio:  1.484                 Gaps:      18
 Percent Similarity:  39.130   Percent Identity: 22.464

Alignment block:
 US-09-126-945B-2 x US-09-344-579-1   ..

 Align seq 1/1  to: US-09-344-579-1  from: 1  to: 2268

        8 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
          |||:::   |||       |||:::|||   ||||||       ||||||||
        5 TCCTCCCCTCCCCTCCACTCGGCCGTCCCTCCTTCCTCCTCCCTCCTCCC 54

       19 oProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaV 46
          ||||       :::   :::          |||:::||||||   |||:::
       55 TCCTCCTCCCGCTCCTGAA.....GAGCGCGCCGCGTGGGGGACGG 95

       46 alGlyLeuGluArqArqAspTrpSerProSerProProAlaThrProGlu 52
          ::   |||                          ||||||   |||   |||
       96 CCCGGTTA.....CTTCCTCCAGACACTGACGAC 124

       53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPhe.....AspMetLeuTy 67
             |||:::::::              |||                :::
      125 TGCGGTGTCGCTCCAGCTCAGAGCTCCCGGAGCCGCCCGGCCAGCGTCCC 174

       67 rProGluAspSerSer.TrpAlaAlaLysAlaPro..............  78
          |||       |||||| |||   |||   :::|||
      175 GCCTCCCTGATCGTCTCTGGCCGGCGCCCTCGCCCTCGCCCGGCGCGCAC 224

       79 ...GlyAlaSerSerArqGluGluProProGluGluProGluGluCysPr 94
             :::|||::::::   |||:::|||||:::::|||:::       ||
      225 CGAGCAGCCGCGGGCGCCGAGCAGCCACCGTCCCCACCAAGCGCCGGCCC 274

       94 oVal.....IleAspSerG 99
          |:::                :::|||:::
      275 TGCCCGCAGCGGCAGGATGAATGATTCGGAATCAAGAATATGGACCAGG 324

       99 luAlaProAlaGly..... 103
            ||||||:::::
      325 TAGCCCCTGTGGCTAACAGTTACAGAGGACACTCAAGCGCCAGCCAGCC 374

      103 ..................... 103
```

```
      375 TTTGACACCTTTGATGGGTCCCTGTTTGCTGTTTTTCCTTCTCTAAATGA 424

      104 .....SerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisS 118
                :::|||:::   ||||||   ||||||          :::|||:
      425 AGAGCAAACACTGCAAGAAGTGCCAACAGGCTTGGATTCCATTTCTCATG 474

      119 er  .....LeuProLeuGlyValGlnSerMetVal 126
                ::                |||         ||||||   |||
      475 ACTCCGCCAACTGTGAATTGCCTTTGTTAACCCCGTGCAGCAAGGCTGTG 524

      127 ValGlyGluValLeuLysAspIleGluThrAlaCysLys..... 149
          :::::::::::|||||          ::::::   |||
      525 ATGAGTCAAGCCTTAAAAGCTACCTTCAGTGGCTTCAAAAAGGAACAGCG 574

      140 .LeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValG 156
          |||:::|||       :::|||     ||||||   :::::|||
      575 GCGCCTGGGCATTCCAAAGAACCCCTGGCTGTGGAGTGAGCAACAGGTAT 624

      156 luLysTrpLeuLeuTrpThrGluHisGlnTyrArqLeuProProMetGly 172
          :::|||||||||||||:::   :::::::::   |||       ::::::
      625 GCCAGTGGCTTCTCTGGGCCACCAATGAGTTCAGTCTGGTGAACGTGAAT 674

      173 LysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGl 189
                     :::   ||:::   ||||||   ::::::::::||
      675 CTGCAGAGGTTCGGCATGAATGGCCAGATGCTGTGTAACCTTGGCAAGGA 724

      189 uGlnPheArqGluArqSerPro...LeuGlyGlyAspValLeuHisAlaH 205
          |:::|||   :::   :::|||   :::   ||||||:::|||      |
      725 ACGCTTTCTGGAGCTGGCACCTGACTTTGTGGGTGACATTTCTCTGGGAAC 774

      205 isLeuAsp..... 207
          |||||:::
      775 ATCTGGAGCAAATGATCAAAGAAAACCAAGAAAAGACAGAAGATCAATAT 824

      208 .............IleTrpLysSerAlaAlaTrpMet..... 215
                      :::   :::         ||:::
      825 GAAGAAAATTCACACCTCACCTCCGTTCCTCATTGGATTAACAGCAATAC 874

      215 ..................................................  215

      875 ATTAGGTTTTGGCACAGAGCAGGCGCCCTATGGAATGCAGACACAGAATT 924

      215 ..................................................  215

      925 ACCCCAAAGGCGCCTCCTGGACAGCATGTGTCCGGCCTCCACACCCAGC 974

      216 .....................................LysGluArqThrSe 220
                                              |||:::|||   ||
      975 GTACTCAGCTCTGAGCAGGACTTCAGATGTTCCCCAAGTCTCGGCTCAG 1024

      220 rProGlyAlaIleHisTyrCysAlaSerThrSerGlu..... 232
          |       ::::::   ||||||:::   :::::::::
     1025 CTCCGTCAGCGTCACCTACTGCTCTGTCAGTCAGGACTTCCCAGGCAGCA 1074

      232 ..................................................  232

     1075 ACTTGAATTTGCTCACCAACAATTCTGGGACTCCCAAAGACCACGACTCC 1124

      233 ..........................................GluSe 234
                                                   :::||
     1125 CCTGAGAACGGTGCGGACAGCTTCGAGAGCTCAGACTCCCTCCTCCAGTC 1174

      234 rTrpThrAsp.....S 238
          |||:::::::      :
     1175 CTGGAACAGCCAGTCGTCCTTGCTGGATGTGCAACGGGTTCCTTCCTTCG 1224

      238 erGluValAspSerSerCysSer..... 245
          :::::   :::::::::|||||
     1225 AGAGCTTCGAAGATGACTGCAGCCAGTCTCTCTGGCTCAATAAGCCAACC 1274

      245 ..................................................  245

     1275 ATGTCTTTCAAGGATTACATCCAAGAGAGGAGTGACCCAGTGGAGCAAGG 1324
```

```
      246 .....GlyGlnP 248
                    ...  |
     1325 CAAACCACTTATACCTGCAGCTGTGCTGGCCCGCTTCACAGGAAGTCCAC 1374

      248 rolleHisLeuTrpGlnPheLeuLysGluLeuLeuLysProHisSer 284
          ||||||:||||||||||||||||   |||   |||||||:::     |||
     1375 CTATTCAGCTGTGGCAGTTTCTCCTGGAG...CTGCTATCAGACAAATCC 1421

      285 TyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIl 281
                ||||||   |||   :::   :::          |||||||:
     1422 TGCCAGTCATTCATCAGCTGGACTGGA...GACGGATGGGAGTTTAAGCT 1468

      281 eGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgP 298
          :   |||      :::||||||||   ||||||  ||||||||||:::|
     1469 CGCCGACCCCGATGAGGTGGCCCGCCGGTGGGGAAAGAGGAAAAATAAGC 1518

      298 roAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLys 314
          ||  ||||||||||:::|||||||||||||:::::|||   ||||||
     1519 CCAAGATGAACTACGAGAAGCTGAGCCGGGGCTTACGCTACTATTACGAC 1568

      315 LysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPh 331
          |||:::|||||:::|||   :::   :::::|||   ||||||:::||
     1569 AACAACATCATCCACAAGACGTCG...GGCAAGCGCTACGTGTACCGCTT 1615

      331 eVal 332
          ||||
     1616 CGTG 1619

seq_name: /cgn2_6/ptodata/1/iia/5A_COMB/us-09-009-913-6

seq_documentation_block:
; Sequence 6, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;    APPLICANT:  AxyS Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  Asthma Related Genes
;    NUMBER OF SEQUENCES:  339
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  SEQ-4P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3231
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-009-913-6

alignment_scores:
            Quality:  299.00              Length:    211
              Ratio:  2.318                 Gaps:      4
 Percent Similarity:  61.137   Percent Identity:  33.175

alignment_block:
 US-09-126-945B-2 x US-09-009-913-6   ..

 Align seq 1/1  to: US-09-009-913-6  from: 1  to: 2428

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
          :::|||||||||              ::::::   |||      ||||:
      333 CAGACAGCCTCTGACTCATACTGGACATCAGTCCACCCTGAATACTGGAC 382

      151 rProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 168
          .      :::|||   :::|||||   :::          |||||:::|
      383 TAAGCGCCATGTGTGGGAGTGGCTCCAGTTCTGCTGCGACCAGTACAAGT 432

      168 euProProMetGlyLysAlaPhe......GlnGluLeuAlaGlyLysGlu 182
          ||              :::|||          ::::::::::|||   :::
      433 TGGACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGGCCTGCAG 482

      183 LeuCysAlaMetSerGlnGluGlnPheArgGlnArgSerProLeuGlyGl 199
          |||||||:::|||:::::|||:::|||   :::   :::   |||   ||
      483 CTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAGCTGGCCTCTGCGG 532

      199 yAspValLeuHisAlaHisLeu..... 206
          |:::   |||:::      |||
      533 CGAGTACCTGTACTTCATCCTCCAGAACATCCGCACACAAGGTTACTCCT 582

      207 ........AspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSer 220
                  |||      :::|||   |||   :::|||:::   ::::::
      583 TTTTTAATCACCCTGAAGAAAGCAAGCCCACCATCAAAGACTATCCTCAT 632

      221 ProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAs 237
          :::            |||   :::||||||        :::      ||
      633 TCCAAC............TGCTTGAAAACAAGTGGCATCAAAAGTCAAGA 670

      237 pSerGlnValAspGerSerCysSerGlyGlnProIleHisLeuTrpGlnP 254
          |   :::      |||      |||   |||      ||||||||||:::|
      671 CTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGGAAT 720

      254 heLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArg 270
          ||:::::::::|||||||||:::|||:::::::          :::::::::
      721 TTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTGGAA 770

      271 TrpLeuAsnLysGlnLysGlyIlePheLysIleGluAspSerAlaGlnVa 287
          |||   ::::::|||:::|||||||||:::::      |||      ::
      771 TGGGAAGATAGGGAACAAGGAATTTTCGGGTGGTTAAATCGGAAGCCCT 820

      287 lAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspL 304
          :|||:::::::|||||   ||||||:::::      |||:::|||:::|
      821 GGCAAAGATGTGGGGACAAAGGAAGAAAAATGACAGAATGACGTATGAAA 870

      304 ysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLys 320
          ||||||||||:::::|||   ||||||||   ||||||:::::::::::
      871 AGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGGAATTTTGGAGCGG 920

      321 ProAspIleSerGlnArgLeuValTyrGlnPhe 331
             |||      :::|||||||||||||:::|||
      921 GTTGAC......CGAAGCTTACTGTACAAATTT 947

Sequence Name: /cgn2_6/ptodata/1/iia/5A_COMB/us-08-344-143B-3

seq_documentation_block:
; Sequence 3, Application US/08344343B
; Patent No. 5968734
;  GENERAL INFORMATION:
;    APPLICANT:  Aurias, Alain
;    APPLICANT:  Delattre, Olivier
;    APPLICANT:  Desmaze, Chantal
```

```
;     APPLICANT:  Meliot, Thomas
;     APPLICANT:  Peter, Martine
;     APPLICANT:  Plouquastel, Beatrice
;     APPLICANT:  Thomas, Gilles
;     APPLICANT:  Zucman, Jessica
;     TITLE OF INVENTION:  NUCLEIC ACID CORRESPONDING TO A GENE OF
;     TITLE OF INVENTION:  CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
;     TITLE OF INVENTION:  TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
;     TITLE OF INVENTION:  TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
;     TITLE OF INVENTION:  TRANSLOCATIONS
;     NUMBER OF SEQUENCES:  129
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Weiser & Associates
;       STREET:  230 South Fifteenth Street
;       CITY:  Philadelphia
;       STATE:  PA
;       COUNTRY:  USA
;       ZIP:  19102
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  AEDIT 1.0 DOS text editor
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/343,443B
;       FILING DATE:  18-NOV 1994
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/FR93/00494
;       FILING DATE:  19-MAY-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  FR 92/06123
;       FILING DATE:  20 MAY-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Weiser, Gerard J.
;       REGISTRATION NUMBER:  19,763
;       REFERENCE/DOCKET NUMBER:  989.6121P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  215 875 8383
;       TELEFAX:  215 875-8394
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2938 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  144..1498
US-08-343-443B-3


alignment scores:
            Quality:  278.50              Length:     882
              Ratio:  1.522                 Gaps:      18
 Percent Similarity:  47.906   Percent Identity: 28.010

alignment block:
 US-09-126-945B-2 x US-08-343-443B-3  ..

 Align seq 1/1  to: US-08-343-443B-3  from: 1  to: 2938

       44 GlyAlaValGlyLeuGluArg.....................Ar 41
          |||:::|||||   ||||||                      ||
      163 GGCTCTGTCGGTGGTGAGCGACGACCAGTCCCTCTTTGACTCAGCGTACG 212

       41 qAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerAlaP 58
          |:::   ||||||||||||               ||||||
      213 GAGCGGCAGCCCATCTCCC.............CAAGGC......... 237

       58 heTyrLeuSerTyrPheAspMetLeufyrProGluAspSerSerTrpAla 74
                          |||  |||
      238 ..........CGACATGACTGCCTC..................... 252

       75 AlaLysAlaProGlyAlaSer...SerArgGluGluProProGluGluPr 90
                  |||  |||      |||  :::::::|
      253 ..........GGGGAGTCCTCACTACGGCAGCCCCACAAGATCAACC 290

       90 oGluGlnCysProVal...IleAspSerGlnAlaProAlaGlySer.... 104
          |          :::   :::|||:::::::|||      |||:::
      291 CCCTCCCACCACAGCAGGAGTGGATCAATCAGCCAGTGAGGGTCAACGTC 340

      105 .........................LeuAspLeuValProGlyGlyLeu 112
                                  ::::::  |||  |||||||||
      341 AAGCGGGAGTATGACCACATGAATGGATCCAGGGAGTCTCCGGTGGACTG 390

      113 .....................................ThrLeuGluGluHi 117
                                             ::::::  ||
      391 CAGCGTTAGCAAATGCAGCAAGCTGCTGGGCGGAGGCGAGTCCAACCCCA 440

      117 sSerLeuGluGlnValGlnSerMetValVal.GlyGluValLeuLysAsp 133
          |:::|||:::|||:::   :::          |||          :::
      441 TGAACTACAACAGCTATATGGACGAGAAGAATGGCCCCCCTCCTCCCAAC 490

      134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTr 150
          :::  |||      :::  :::  :::  |||||||||       ||
      491 ATGACCACCAACGAGAGGAGAGTCATCGTCCCCGCAGACCCCACACTGTG 540

      150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrA 167
          |:::   :::::|||:::::|||||   |||:::      :::|||
      541 GACACAGGAGCATGTGAGGCAATGGCTGGAGTGGGCCATAAAGGAGTATA 590

      167 rqLeuProProMetGlyLysAla...PheGlnGluLeuAlaGlyLysGlu 182
            |||   :::   :::  |||||||:::::  |||||||||
      591 GCTTGATGGAGATCGACACATCCTTTTCCAGAACATGGATGGCAAGGAA 640

      183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeu...Gl 198
          ||||||   |||:::::|||:::|||   :::   :::   |||   ::
      641 CTGTGTAAAATGAACAAGGAGGACTTCCTCCGCGCCACCACCCTCTACAA 690

      198 yGlyAspValLeuHisAlaHisLeu..................... 206
          :   :::||||||   :::||||||
      691 CACGGAAGTGCTGTTGTCACACCTCAGTTACCTCAGGGAAAGTTCACTGC 740

      206 ..................................................... 206

      741 TGGCCTATAATACAACCTCCCACACCGACCAATCCTCACGATTGAGTGTC 790

      207 ..................AspIleTrpLysSerAlaAlaTrpMetLysGl 217
                             |||     :::   :::||||||   :::::
      791 AAAGAAGACCCTTCTTATGACTCAGTCAGAAGAGGAGCATGGGGCAATAA 840

      217 nArqThrSerPro.....................GlyAlaIleHisTyrC 227
          :   :::|||                      ||||::   :::
      841 CATGAATTCTGGCCTCAACAAAAGTCCTCCCCTTGGAGGGCACAAACGA 890

      227 ysAlaSerThrSerGluGlu............................. 233
            ::::::::::::|||:::
      891 TCAGTAACAATACAGAGCAACGGCCCCAGCCAGATCCGTATCAGATCCTG 940

      234 SerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHi 250
          :::   |||:::|||::::::   :::      |||||||||   ||||:
      941 GGCCCGACCAGCAGTCGCCTAGCCAACCCTGGAAGCGGGCAG...ATCCA 987

      250 stLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyA 267
          :|||||||||||||||   |||   ||||||:::      :::   :::
      988 GCTGTGGCAATTCCTCCTGGAG...CTGCTCTCCGACAGCGCCAACGCCA 1034

      267 rqPheIleArqTrpLeuAsnLysGluLysGlyIlePheLysIleGluAsp 283
            |||  |||  :::      ::::||   ||||||:::   |||
     1035 GCTGTATCACCTGG...GAGGGACCAACGGGGAGTTCAAAATGACGGAC 1081

      284 SerAlaGluValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMe 300
          :::|||||||||   ||||||   ||||||::::::|||   ||
     1082 CCCGATGAGGTGGCCAGGCGCTGGGGCGAGCGGAAAAGCAAGCCCAACAT 1131
```

```
      300 tAsnIyrAspLysLeuSerArqSerIleArqGlnTyrIyrLysLysGlyI 347
          ||||||||||||||||||||||||||||:::::|||   ||||||   |||:::|
     1132 GAATTACGACAAGCTGAGCCGGGCCCTCCGTTATTACTATGATAAAAACA 1181

      317 IeIIeArqLysProAspIleSerGlnArqLeuValIyrGluPhe 331
          |:::.   |||          :::::|||   :::|||:::|||
     1182 TTATGACCAAAGTGCAC...GGCAAAAGATATGCTTACAAATTT 1222

seq_name: /cgn2_6/ptodata/1/pubseq/PCTUS_NEW_SEQ/US-09-009-913-4;

seq_documentation_block:
; Sequence 4, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;     APPLICANT:  AxyS Pharmaceuticals, Inc.
;     TITLE OF INVENTION:  Asthma Related Genes
;     NUMBER OF SEQUENCES:  339
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Bozicevic & Reed, LLP
;       STREET:  285 Hamilton Ave, Suite 200
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94301
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/009,913
;       FILING DATE:  21-JAN-1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sherwood, Pamela J
;       REGISTRATION NUMBER:  36,677
;       REFERENCE/DOCKET NUMBER:  SEQ 4P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-327-3241
;       TELEFAX:  650-327-3241
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5667 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-09-009-913-4


alignment_scores:
             Quality:  266.00              Length:     428
               Ratio:  1.298                Gaps:      18
  Percent Similarity:  47.897   Percent Identity:  25.467

alignment_block:
 US-09-126-945B-2 x US-09-009-913-4   ..

 Align seq 1/1  to: US-09-009-913-4  from: 1  to: 5667

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
          :::|||:::   :::        :::   ||||||   :::|||:::||
       57 ITGGGAGGCATTGGCTCCACTCCAGCCTGGGTGTCTACAGACCACATTCT 106

       17 q...LeuProProAspIhrValSerArqThr.GlyLeuGluLysAlaAla 42
          |   ||| ||          :::|||||| |||          :::
      107 CACCCTGCCTTTGTTACTGGGAAACCGAACGCGGCGCTGTGGCTTTCAGC 156

       33 AlaGlyAlaValGlyLeuGlu.........................Ar 40
```

```
          |||      ||||||:::                            ||
      157 TTGGGTAAGCCGGGTCTGCGGCGGGGATTGCCATCTGAAGACAGAGGCAG 206

       40 qArqAspTrpSerProSerProProAlaThrPro.............. 51
          ||||      :::|||   |||   |o|   |||
      207 GAGGGCAGCAACACCTTGCCCAGCTGCACACCCAGTAACAAGTTTCCTCA 256

       52 ..............................GluGlnGlyLeuSerAla 57
                                        :::||||||||||
      257 GTGCGGGTATCTGCCACAGGCTGGGCTCGTCATCAAAGGGCCTCAGTCAT 306

       58 PheTyrLeuSerTyrPheAspMetLeu.....TyrProGluAs 70
          :::   ::::::   |||      |||          |||   |||
      307 ATTTTAATAGAGCTCTTCAAGTATCTGGCTTTGTGATAATATCAGGAA.. 354

       70 pSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArqGluGlnProP 87
                          :::   |||   |||:::|||::
      355 ..................TCAGTTGGTTTCTCTGACAGACACTG..... 380

       87 roGluGluProGluGlnCysProValIleAspSerGlnAlaProAlaGly 103
                             |||:::|||                :::|||
      381 ..................CCCATTATCATGATTCTGGAAGGAGGTGGT 410

      104 SerLeuAspLeuValProGlyGlyLeuThrLeu.................. 114
          :::::|||   ||||||:::      |||
      411 GTAATGAATCTCAACCCGGGCAACAACCTCCTTCACCAGCCGCCAGCCTG 460

      115 ....GluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluV 130
              ::::::::|||      :::|||:::|||          |||
      461 GACAGACAGCIACTCCACGTGCAATGTTTCCAGTGGGTITTTTGGAGGCC 510

      130 alLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAsp 146
                ::::|||:::
      511 AGTGGCATGAAATTCAT................................ 527

      147 ProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGl 163
          |||:::   |||:::      :::|||   :::|||||
      528 CCTCAGTACTGGACCAACTACCAGGTGTGGGAGTGGCTCCAGCACCTCCT 577

      163 .HisGlnTyrArgLeuProMetLysGlyLysAlaProGlnGlu......L 178
                    :::|||                |||||||||          :
      578 GGACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTTCCACA 627

      178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArqGlnArq 194
          ::   |||:::::|||||:::|||||   :::::|||   :::
      628 TCAACGGCGAGCACCTCTCCAGCATGAGTTTCAGGAGTTCACCCGGCGG 677

      195 SerProLeuGlyGlyAspValLeuHisAlaHisLeu.............. 206
          :::      :::|||:::::::|||:::::::|||
      678 GCAGGCACCGCGGGGCAGCTCCTCTACAGCAACTTGCAGCATCTGAAGTG 727

      207 ...................AspIleIrpLysSer.........AlaAlaI 214
                             |||:::::::::::||          :::
      728 GAACGGCCAGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGTCATTG 777

      214 rpMetLysGluArgThrSerProGlyAlaIle................... 224
                |||:::|||:::|||:::   :::
      778 TCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAGACGAG 827

      225 HisTyrCysAlaSerThrSerGluGluSerIrpThrAspSerGluValAs 241
          :::|||      :::|||:::      |||   :::|||      :::||
      828 AACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG...TTGGA 874

      241 pSerSer......CysSerGlyGln....................... 247
          ||||:::      |||   :::|||
      875 CAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAGTCACC 924

      248 ..ProIle........................................ 249
            ||||::
      925 TTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCCCTGCC 974

      250 ..................................HisLeuTrpGlnPh 264
                                           ||||||||||:::||
```

```
      975 AAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGCAATT 1024

      254 cLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgI 271
          |:::::::::::::|||||::|||   :::          :::||||:|
     1025 CATCTGCACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATAAAAT 1074

      271 rpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVal 287
          ||   :::::::::::|||:::|||::::::      |||      |||
     1075 GGCAAGACCGATCTGAGGCCGTCTTCAGGTTCTTGAAATCAGAGGCAGTG 1124

      288 AlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLy 304
          |||:::|||||||||   :::|||||:::   :::|||:::|||:::||
     1125 GCTCAGCTATGGGGTAAAAAGAAGAACAACACCAGCATGAACTATGAAAA 1174

      304 sLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysP 321
          ||||||||||:::::|||   ||||||||||:::   |||:::::::::
     1175 GCTCAGCGGAGCTATGAGATATTACTACAAAAGAGAAATACTGGAGCGTG 1224

      321 roAspIleSerGlnArgLeuValTyrGlnPhe 341
            |||   ::::::|||||||||||::::||
     1225 TGCAT...GGACGAAGACTGGTATATAAATTT 1253

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB/ascii/US08_868_281_3

seq_documentation block:
; Sequence 3, Application US/08868281
; Patent No. 5721113
;  GENERAL INFORMATION:
;     APPLICANT:  Libermann, Towia A
;     APPLICANT:  Oettgen, Joerg P
;     APPLICANT:  Kunsch, Charles A
;     TITLE OF INVENTION:  NERF Genes
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporation
;       STREET:  709 Swedeland Road
;       CITY:  King of Prussia
;       STATE:  Pennsylvania
;       COUNTRY:  USA
;       ZIP:  19406
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/868,281
;       FILING DATE:
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bannister, Kirk
;       REGISTRATION NUMBER:  34,833
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  610 270 5096
;       TELEFAX:  610 270 5090
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4240 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US 08 868 281 3


alignment scores:
             Quality:  264.00              Length:     845
               Ratio:  1.584                Gaps:      11
 Percent Similarity:  48.116   Percent Identity: 24.058

alignment block:
 US 09 126 945B 2 x US 08 868 281 3   ..
```

```
Align seq 1/1  to: US 08 868 281 3  from: 1  to: 4240

       20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 36
          ||::::   ::::::   ||::::             :::::|||::::::
      192 CCTCAGGGAATAACAATGACATCAGCAGTGGTTGACAGTGGAGGTACTAT 241

       36 I.................GlyLeuGluArgArgAspTrpSer..... 44
          :                 |||::::|||:::::::::   |||
      242 TTTGGAGCTTTCCAGCAATGGAGTAGAAAATCAAGAGGAAAGTGATAAGG 291

       45 ............................ProSerProProAlaThrPro 51
                                       |||   |||   |||
      292 TTTCTGAATATCCAGCAGTGATTGTGGAGCCAGTTCCAAGTGCCAGATTA 341

       52 GluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyr.. 67
          |||||||||   ::::|||      |||::::|||   |||      |||
      342 GAGCAGGGGTATGCAGCCCAGTTCTGGGTTATGATGATGACATTATTT 391

       68 .ProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
          :::|||   :::                                     |
      392 GATGCAAGATGTGCCA.................................G 408

       84 IuGluProProGluGluProGluGlnCysProValIleAspSer..... 98
          |||||         |||   |||:::::   ::::::::::::
      409 AAGAACAAGAAGTTGAGACCGACAATGCGGAAACACTGGAAGCATCAGTT 458

       99 .....................................GlnAlaProAlaGl 103
                                              :::|||   :::
      459 CACAGCAGTAATGCACACTGTACAGATAAGACAATTGAAGCTGCTGAAGC 508

      104 ySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuG 120
          :   |||   :::                 ::::::::::||   |
      509 CCTGCTTCATATGGAATCTCCTACCTGCTTGAGGGATTCAAGAAGTCCTG 558

      120 IuGlnValGluSerMetValValGlyGluValLeuLysAspIleGluThr 136
          ||   :::::::::   :::      ::|||:::          ||||||
      559 AATTCATCCATGCTGCTATGAGGCCAGATGTCATT......ACAGAAACT 602

      137 AlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSe 153
          :::   ::::::   :::      ::::::||||||||   ||||||
      603 GTAGTGGAGGTGTCAACTGAAGAGTCTGAACCCATGGATACCTCTCCT.. 650

      153 rAsnValGlnLysTrpLeuLeuTrpThrGluHisGluTyrArgLeuProP 170
                                                 ::::||
      651 .............................................ATTCCAA 657

      171 roMetGlyLysAlaArgLeuIleLeuLeuAlaLeuLysLysLeuCysAlaMet 186
                ::: ::: ::: ::: ||:::      :::
      658 CATCACCAGATAGCCATGAACCAATGAAAAAGAAAAAA.....GTT 698

      187 SerGluGluGlnPheArgGluArgSerProLeuGlyGlyAspValLeuHi 203
          :::::::::      ||::::||||||:::::::::
      699 GGCCGTAAACCAAAGACCCAGCAATCACCAATTTCCAATGGC..... 740

      204 sAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrS 220
                          :::              :::|||::::::
      741 ...................tCTCCTGAGTTAGGTATAAAGAAGAAACCCAA 771

      220 erProGlyAlaIleHistyrCysAlaSerThrSerGluGluSerTrpThr 236
          |||
      772 GAGAAGGA.......................................... 779

      237 AspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGl 253
                                 ::::|::::      ::::::::::
      780 ........................AAAGGAAACACAACCTATTTGTGGGA 805

      254 nPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleA 270
          :||||||   :::||||||         ::::::      |||:::|||:
      806 GTTTCTTTTAGATCTACTTCAAGATAAAAATACTTGTCCCAGGTATATTA 855

      270 rgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlu 286
          ::|||   :::::|||||||||||||||||:::   ||||||
      856 AATGGACTCAGAGAGAAAAAGGCATATTCAAGCTGGTGGATTCAAAGGCT 905
```

```
      287 ValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAs 303
          |||:::::::|||||||||   :::|||||||:::|||   |||||||||:::
      906 GTCTCTAAGCTTTGGGGAAAGCATAAGAACAAACCAGACATGAACTATGA 955

      304 pLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgL 320
          :   :::::::|||:::::::|||   |||||||:::::|||||||:::   |
      956 AACCATGGGACGAGCTTTGACATACTACTACCAAAGGGGAATTCTTGCAA 1005

      320 ysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          ||   :::   :::|||||||||||||||||||||
     1006 AGGTTGAA...GGACAGAGGCTTGTATATCAGTTC 1037

Seq_name: Query_2 ; Length 5427 ; MB Seq:US-09-009-913-2

Seq_documentation_block:
; Sequence 2, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;    APPLICANT:  Avys Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  Asthma Related Genes
;    NUMBER OF SEQUENCES:  339
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  SEQ 4P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3231
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5427 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-009-913-2


alignment_scores:
            Quality:  262.00              Length:     463
              Ratio:  1.424                 Gaps:      18
 Percent Similarity:  50.689   Percent Identity: 25.895

alignment_block:
 US-09-126-945B-2 x US-09-009-913-2  ..

 Align seq 1/1  to: US-09-009-913-2  from: 1  to: 5427

       41 ArgAspTrpSerProSerProProAlaThrProGluGlnGly..... 54
          |||  |||  |||  |||  ::::::  |||  :::
       88 CGCCCGTGGTGCCCCATCCCTATAGGAGCTGGTGAGATTGCAGCCTGCTG 87

       55 .LeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGlu.... 69
           |||  ::::::     |||||:::  :::  :::|||||
       88 CCTCCCCTCCATCAGCCACAGCTATTGG...ATTTCCCACCCAGAATCTT 134

       70 ................AspSerSer......TrpAlaAlaLysAlaPro 78
                          |||||||::     |||:::  :::
      135 TAGGTAAATCAGATCATGATTCTGGAAGGAGGTGGTGTAATGAA...... 178

       79 GlyAlaSerSerArgGluGluProProGluGluProGluGlnCysProVa 95
          ::::::  |||:::::::|||||
      179 ...TCTCAACCCCGGCAACAACCTCCT..................... 202

       95 IleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyL 112
                  :::|||||:::::|||||||
      203 ............TCACCAGCCGCCAGCCTGGAC.............. 223

      112 euThrLeuGluGlu.HisSerLeuGluGlnValGlnSerMetValValGl 128
             :::::: :::|||      :::|||:::|||         ||
      224 ........AGACAGCTACTCCACGTGCAATGTTTCCAGTGGGTTTTTGG 265

      129 yGluValLeuLysAspIleGlyThrAlaCysLysLeuLeuAsnIleThrA 145
          |              :::|||:::
      266 AGGCCAGTGGCATGAAATTCAT........................... 287

      145 laAspProMetAspTrpSerProSerAsnValGluLysTrpLeuLeuTrp 161
           |||:::   |||:::      :::|||   :::|||||||
      288 .....CCTCAGTACTGGACCAAGTACCAGGTGTGGGAGTGGCTCCAGCAC 332

      162 ThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu.. 177
                     :::|||                        |||||||||
      333 CTCCTGGACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTT 382

      178 ....LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgG 193
              :::   |||:::::::|||||:::|||||   :::::|||   :
      383 CGACATCAACGGCCAGCACCTCTGCAGCATGAGTTTGCAGGAGTTCACCC 432

      193 lnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu......... 206
          ::   :::      :::|||:::::::|||:::::::|||
      433 GGGCGGCAGGCACGGCGGGGCAGCTGCTCTACAGCAACTTGCAGCATCTG 482

      207 ........................AspIleTrpLysSer.........Al 212
                                  |||:::::::::|||       ::
      483 AAGTGGAACGGCCAGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGT 532

      212 aAlaTrpMetLysGluArgThrSerProGlyAlaIle.............. 224
          :          |||:::|||:::|||:::   :::
      533 CATTGTCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAG 582

      225 .....HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGlu 239
           :::|||      :::|||:::      |||  :::|||
      583 ACGAGAACTATTTATATGACACCAACTATCGTAGCACAGTAGATTTG    629

      240 ValAspSerSer      CysSerGlyGln    .                 247
          :::|||||:::      |||   :::|||
      630 TTGGACAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAG 679

      248 .......ProIle..................................... 249
                 |||:::
      680 TCACCTTCCTGTTGCAGAGTCACCTGATATCAAAAAGGAGCAAGACCCCC 729

      250 .....................................HisLeuTrp 252
                                              |||||||||
      730 CTGCCAAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGG 779

      253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPhe11 269
          :::|||:::::::::::|||||:::|||   :::          :::||
      780 GAATTCATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAAT 829

      269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaG 286
          |:::|||   :::::::::::|||:::|||:::::      |||
      830 AAAATGGGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGG 879

      286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 402
```

```
          ||||||:::|||||||||   :::|||||||:::   :::|||:::|||
      880 CAGTGGCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTAT 929

      303 AspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 319
          :::||||||||||||:::::|||   ||||||||||:::   |||:::::
      930 GAAAAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATACTGGA 979

      319 qLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          ::::   |||   ::::::|||||||||||||:::|||
      980 GCGTGTGGAT...GGACGAAGACTGGTATATAAATTT 1013

Seq name: /cgn2_6/ptodata/1/zinq/26A_COMB.seq:US_09_009_913_4

Seq_documentation block:
; Sequence 4, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;     APPLICANT:  AxyS Pharmaceuticals, Inc.
;     TITLE OF INVENTION:  Asthma Related Genes
;     NUMBER OF SEQUENCES:  339
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Bozicevic & Reed, LLP
;       STREET:  285 Hamilton Ave, Suite 200
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94301
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/009,913
;       FILING DATE:  21 JAN 1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sherwood, Pamela J
;       REGISTRATION NUMBER:  36,677
;       REFERENCE/DOCKET NUMBER:  SEQ 4P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650 327 3241
;       TELEFAX:  650 327 3231
;       TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5510 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US 09 009 913 4


Alignment scores:
            Quality:  262.00              Length:     329
              Ratio:  1.497                 Gaps:      13
 Percent Similarity:  53.191   Percent Identity: 25.228

Alignment block:
 US 09 126 945B 2 x US 09 009 913 4  ..

 Align seq 1/1  to: US 09 009 913 4  from: 1  to: 5510

       71 SerSerTrpAlaAlaLysAlaProGlyAlaSerSer.....ArgGluG 85
          |||:::|||:::::::::::::   |||::::::::      :::   ||
      152 TCAGCTTGGGTAAGCCGGGTCTGCGGCGGGATTGCCATCTGAAGACAGA 201

       85 nProProGluGluProGluGluCysProValIleAspSerGlnAlaProA 102
          |         :::|||:::   :::   ::::::      ::::::   :

      202 GGCAGGAGGGCAGCCACACCTTGCCCAGATCATGATTCTGGAAGGA...G 248

      102 taGlySerLeuAspLeuValProGlyGlyLeuThrLeu..... 114
          ::|||   ::::::|||   ||||||:::       |||
      249 GTGGTGTAATGAATCTCAACCCCGGCAACAACCTCCTTCACCAGCCGCA 298

      115 .........GluGluHisSerLeuGlnGlnValGlnSerMetValValGl 128
                  :::::::::|||      :::|||:::|||          ||
      299 GCCTGGACAGACAGCTACTCCACGTGCAATGTTTCCAGTGGGTTTTTTGG 348

      128 yGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrA 145
          |           :::|||:::
      349 AGGCCAGTGGCATGAAATTCAT..........................  370

      145 laAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrp 161
               ||:::   |||:::      :::|||   :::|||||||
      371 .....CCTCAGTACTGGACCAAGTACCAGGTGTGGGAGTGGCTCCAGCAC 415

      162 ThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu.. 177
                     :::|||                  |||||||||
      416 CTCCTGGACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTT 465

      178 ....LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgG 193
              :::   |||:::::::|||||||:::||||||   :::::::|||   :
      466 CGACATCAACGGCGAGCACCTCTGCAGCATGAGTTTGCAGGAGTTCACCC 515

      193 lnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu..... 206
          ::   :::      :::|||:::::::|||:::::::::|||
      516 GGCGGGAGGGAGGGCGGGCAGCTCCTCTACAGCAACTTGCAGCATCTG 565

      207 .........................AspIleTrpLysSer.........Al 212
                                  |||:::::::::|||         ::
      566 AAGTGGAACGGCCAGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGT 615

      212 aAlaTrpMetLysGlnArgThrSerProGlyAlaIle..... 224
          :           |||:::|||:::|||:::   :::
      616 CATTGTCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAG 665

      225 .....HistyrCysAlaSerThrSerGluGluSerTrpThrAspSerGlu 239
              :::|||      :::|||:::      |||   :::|||
      666 ACGAGAACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG... 712

      240 ValAspSerSer.....CysSerGlyGln..... 247
          :::|||||||::      |||   :::|||
      713 TTGGACAGCAAAACTTCTGCCGGGCTCAGATCTCCATGACAACCACCAG 762

      248 .......ProIle..................................... 249
                ||||::
      763 TCACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCC 812

      250 .....................................HisLeuTrp 252
                                              |||||||||
      813 CTGCCAAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGG 862

      253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl 269
          :::|||:::::::::::::|||||::::|||   :::         :::|||
      863 GAATTCATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAAT 912

      269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaC 286
          |:::|||   :::::::::::::|||:::|||::::::      |||
      913 AAAATGGGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGG 962

      286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
            ||||||:::|||||||||   :::|||||||:::   :::|||:::|||
      963 CAGTGGCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTAT 1012

      303 AspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 319
          :::|||||||||||||:::::|||   ||||||||||:::   |||:::::
     1013 GAAAAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATACTGGA 1062

      319 qLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          ::::   |||   ::::::|||||||||||||:::|||
     1063 GCGTGTGGAT...GGACGAAGACTGGTATATAAATTT 1096
```

```
Seq_name: /cgn2_6/prodata/1/ina/5A_COMB_seq/US-08-306-691B-43

seq_documentation_block:
; Sequence 43, Application US/08306691B
; Patent No. 5734039
;  GENERAL INFORMATION:
;     APPLICANT:  Calabretta, Bruno
;     APPLICANT:  Skorski, Tomasz
;     TITLE OF INVENTION:  ANTISENSE
;     TITLE OF INVENTION:  OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Seidel, Gonda, Lavorgna & Monaco, P.C.
;       STREET:  Two Penn Center, Suite 1800
;       CITY:  Philadelphia
;       STATE:  Pennsylvania
;       COUNTRY:  U.S.A.
;       ZIP:  19102
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 720 Kb
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  WordPerfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  08/306,691B
;       FILING DATE:  September 15, 1994
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Monaco, Daniel A.
;       REGISTRATION NUMBER:  30,480
;       REFERENCE/DOCKET NUMBER:  8321-8
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (215) 568-8383
;       TELEFAX:  (215) 568-5549
;       TELEX:  No. 5734039e
;   INFORMATION FOR SEQ ID NO:  43:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1604 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-306-691B-43


alignment_scores:
            Quality:  251.50              Length:    530
              Ratio:  1.251                 Gaps:     22
 Percent Similarity:  37.925   Percent Identity: 21.698

alignment_block:
 US-09-126-945B-2 x US-08-306-691B-43   ..

 Align seq 1/1  to: US-08-306-691B-43  from: 1  to: 1604

       24 SerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluAr 40
          ::::|||::|||   |||:::   |||   |||:::
       10 GCGAGGGCCGGGCACGAGGAGCGGGGCGGCGCGGCGGG.............. 45

       40 qArqAspIrpSerProSerProProAlaThrProGluGlnGlyLeuSerA 57
           |||   |||:::|||||      :::::|||:::::::::
       46 .CGAGGCTGGGACCCGAGCGGCGCTCACTTCGCCGCAAAGTGCC....... 87

       57 laPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrp 73
                                                  :::      |||
       88 .....................................AACTTCCCCTGG 99

       74 AlaAla...LysAlaProGly..AlaSerSerArqGluGluPro..... 86
          :::|||   :::|||||:::  |||   :::::|||:::
      100 AGTGCCGGCGCCGCACCGTCCGGGCGCGGGGAAAGAAAGGCAGCGGGAA 149

       87                                  ProGlnGluProGlnGlnGy 93
                                           |||   :::|||:::   ||
      150 TTTGAGATTTTTGGGAAGAAAGTCGGATTTCCCCCGTCCCCTTCCCCCTG 199

       93 s.....................................................P 94
          |                                                     |
      200 TTACTAATCCTCATTAAAAAGAAAAACAACAATAACTGCAAACTTGCTAC 249

       94 roValIleAspSerGlnAlaProAla.........GlySerLeuAspLeu 107
          ||         |||   :::|||:::         :::::::::||||||
      250 CAICCCGTACGICCCCCACICCTGGCACCATGAAGGCGGCCGTCGATCTC 299

      108 ValPro.................................................. 109
          |||
      300 AAGCCGACTCTCACCATCATCAAGACGGAAAAAGTCGATCTGGAGCTTTT 349

      110 ......................GlyGlyLeuThrLeuGluGluHisSerL 119
                                :::   :::   |||         |||
      350 CCCCTCCCCGGATATGGAATGTGCAGATGTCCCACTATTAACTCCAAGCA 399

      119 qqGlnGlnV   :::nSerMetValValGlyGluValLeuLysAspIleGlu 135
           ::::::::::::|||:::::::::      :::::::::   :::
      400 GCAAAGAAATGATGTCTCAAGCATTAAAAGCTACTTTCAGTGGTTTCACT 449

      136 ThrAlaCysLysLeuLeuAsn|:hrAlaAspProMetAspIrpSerSr 152
                :::   |||:::|||      ||||||   :::|||:::
      450 AAAGAACAGCAACGACTGGGGATCCCAAAAGACCCCCGGCAGTGGACAGA 499

      152 oSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArqLeuP 169
           ::::::|||:::   |||:::::|||:::   :::::::::   |||
      500 AACCCATGTTCGGGACTGGGTGATGTGGGCTGTGAATGAATTCAGCCTG. 548

      169 roProMetGlyLysAlaPheGlnGlu......LeuAlaGlyLysGluLeu 183
                |||      ||||||:::      :::   |||      |||
      549 .....AAAGGTGTAGACTTCCAGAAGTTCTCTATGAATGGAGCAGCCCTC 598

      184 CysAlaMetSerGluGluGlnPheArqGluArqSerPro...LeuGlyGl 199
          ||||||:::::::::::   |||   :::   :::|||   :::   ||
      594 TGCGCCCTGGGTAAAGACTGCTTTCTCGAGCTGGCCCCAGACTTTGTTGG 643

      199 yAspValLeuHisAlaHisLeuAspIleTrpLys..... 210
          ||||:::|||      |||||||:::|||   :::
      644 GGACATCTTATGGGAACATCTAGATCTCTGCAGAAAGAGGATGTGAAAC 693

      211 .....................SerAlaAlaIrpMetLysGluArq...Thr 219
                               :::   |||:::   :::::::|||   |||
      694 CATATCAAGTTAATGGAGTCAACCCAGCCTATCCAGAATCCCGCTATACC 743

      220 SerProGlyAlaIleHisTyr.................CysAlaSerTh 230
          |||          |||   |||                |||:::
      744 TCCGATTACTTCATTAGCTATGGTATTGAGCATGCCCACTCTGTTCCACC 793

      230 rSerGlu.......................GluSerTrp.......... 245
           ||||||                     |||||||:::
      794 ATCGGAGTTCTCAGAGCCCAGCTTCATCACAGAGTCCTATCAGACGCTCC 843

      235 ..................................................... 235

      844 ATCCCATCAGCTCGGAAGAGCTCCTCTCCCTCAAGTATCAGAATGACTAC 893

      236 ............................ThrAspSer............ 248
                                       |||||||:::
      894 CCCTCGGTCATTCTCCGAGACCCTCTCCAGACAGACACCTTGCAGAATGA 943

      239 ...................GluVal.........AspSerSerCysSerG 246
                             ||||||         |||:::   |||   |
      944 CTACTTTGCTATCAAACAAGAAGTCGTCACCCCAGACAACATGTGCATGG 993

      246 lyGln.................................................. 247
          ||:::
      994 GGAGGACCAGTCGTGGTAAACTCGGGGGCCAGGACTCTTTTGAAAGCATA 1043
```

```
     247 ..................................................  247

    1044 CAGAGCTACCATAGTTGTGATCGCCTCACCCAGTCCTGGAGCAGCCAGTC  1093

     247 ..................................................  247

    1094 ATCTTTCAACAGCCTGCAGCGTGTTCCCTCCTATGACAGCTTCGACTCAG  1143

     247 ..................................................  247

    1144 AGGACTATCCGGCTGCCCTGCCCAACCACAAGCCCAAGGGCACCTTCAAG  1193

     247 ..................................................  247

    1194 GACTATGTGCGGGACCGTGCTGACCTCAATAAGGACAAGCCTGTCATTCC  1243

     248 ..................................ProIleHisLeuTrpG  253
                                           ||||||:::|||||||
    1244 TGCTGCTGCCCTAGCTGGCTACACAGGCAGTGGACCAATCCAGCTATGGC  1293

     253 InPheLeuLysGluLeuLeuLysProHisSerTyrGlyArqPheIle  269
          ||||||||   |||   ||||||          |||          ||||||
    1294 AGTTTCTTCTGGAA...TTACTCACTGATAAATCCTGTCAGTCTTTTATC  1340

     270 ArqTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl  286
          |||   :::   :::          ||||||:::::|||   :::
    1341 AGCTGGACAGGA...GATGGCTGGGAATTCAAACTTTCTGACCCACATGA  1387

     286 nValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMetAsnTyrA  303
          :||||||||||   ||||||   ||||||||||:::|||   ||||||||||:
    1388 CGTGGCCAGGAGATGGGAAAGAGGAAAAACAAACCTAAGATGAATTATG  1437

     303 splysLeuSerArqSerIleArqGlnTyrTyrLysLysGlyIleIleArq  319
          ::|||||||||||||:::::|||   ||||||   ||||:::||||||:::
    1438 AGAAACTGAGCCGTGGCCTACGCTACTATTACGACAAAAACATCATCCAC  1487

     320 LysProAspIleSerGlnArqLeuValTyrGlnPheVal  332
          |||          ::::::|||   ||||||:::||||||
    1488 AAG...ACAGCGGGGAAACGCTACGTGTACCGCTTTGTG  1523

seq name: /seqa/*/pct/dtd/2/ind/PCTUS/COMB.seq:PCT-US93-06251-9

seq documentation block:
; Sequence 9, Application PC/TUS930625I
;  GENERAL INFORMATION:
;     APPLICANT:  Wickstrom, Eric and Rife, Jason P.
;     TITLE OF INVENTION:  Trivalent Synthesis of Oligonucleotides Containing
;     TITLE OF INVENTION:  Stereospecific Alkylphosphonates and Arylphosphonates
;     NUMBER OF SEQUENCES:  93
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;       STREET:  400 Garden City Plaza
;       CITY:  Garden City
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  11530
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/06251
;       FILING DATE:  19930630
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  DiGiglio, Frank S.
;       REGISTRATION NUMBER:  31,346
;       REFERENCE/DOCKET NUMBER:  8586
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  516 742 4343
;       TELEFAX:  516 742 4366
;       TELEX:  230 901 SANS UR
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1604 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
PCT-US93-06251-9


alignment scores:
            Quality:  251.50              Length:    530
              Ratio:  1.251                 Gaps:     22
 Percent Similarity:  37.925   Percent Identity: 21.698

alignment_block:
 US-09-126-945B-2 x PCT-US93-06251-9   ..

 Align seq 1/1  to: PCT-US93-06251-9  from: 1  to: 1604

      24 SerArqThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluAr  40
         :::|||:::|||   ||||::   |||   ||||::
      10 GCGAGGGCCGGGCAGGAGGAGCGGGCGCGCGGCGGG.............  45

      40 qArqAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerA  57
          |||   |||:::||||||      :::::|||:::::::::
      46 .CGAGGCTGGGACCCGAGCCGCCTCACTTCGCCGCAAAGTGCC.......  87

      57 laPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrp  73
                                                  :::      |||
      88 .....................................AACTTCCCCTGG  99

      74 AlaAla...LysAlaProGly..AlaSerSerArqGluGluPro......  86
         :::|||   :::||||||||::  |||   ::::::|||:::
     100 AGTGCCGGGCGCGCACCGTCCGGGCGCCGGGGCAAAGAAAGGCAGCGGGAA 149

      87 ..............................ProGluGluProGluGlnCy  93
                                        |||   ::||||:::   ||
     150 TTTGAGATTTTTGGCAAGAAAGTCGGATTTCCCCCGTCCCCTTCCCCCTG 199

      93 s.............................................P  94
         |                                             |
     200 TTACTAATCCTCATTAAAAACAAAAACAACAATAACTGCAAACTTGCTAC 249

      94 roValIleAspSerGlnAlaProAla.........GlySerLeuAspLeu 107
         ||          |||   :::|||:::          :::::::|||||
     250 CATCCCGTACGTCCCCCACTCCTGGCACCATGAAGGCGGCCGTCGATCTC 299

     108 ValPro..................................................  109
         |||
     300 AAGCCGACTCTCACCATCATCAAGACGGAAAAAGTCGATCTGGAGCTTTT  349

     110 ......................GlyGlyLeuThrLeuGluGluHisSerL  119
                               :::   :::   |||         |||
     350 CCCCTCCCCGGATATGGAATGTGCAGATGTCCCACTATTAACTCCAAGCA  399

     119 euGluGlnValGlnSerMetValValGlyGluValLeuLysAspIleGlu  135
          ::::::::::|||:::::::::     :::::::   :::
     400 GCAAAGAAATGATGTCTCAAGCATTAAAAGCTACTTTCAGTGGTTTCACT  449

     136 ThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerPr  152
              :::   |||:::|||      ||||||   :::|||:::
     450 AAAGAACAGCAACGACTGGGGATCCCAAAAGACCCCCGGCAGTGGACAGA  499

     152 oSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArqLeuP  169
          ::::::|||:::  |||:::::::|||:::  :::::::::  |||
     500 AACCCATGTTCGGCACTGGGTGATGTGGCTGTGAATGAATTCAGCCTG.  548

     169 roProMetGlyLysAlaPheGlnGla.....LeuAlaGlyLysGluLeu  183
                |||      ||||||:::      :::  |||      |||
     549 .....AAAGGTGTAGACTTCCAGAAGTTCTGTATGAATGGAGCAGCCCTC  593

     184 CysAlaMetSerGluGluGlnPheArqGlnArqSerPro...LeuGlyGl  199
```

```
     ||||||||||||||||   |||   :::   :::|||   :::   ||
 594 TGCCCCCTGGGTAAAGACTGCTTTCTCGAGCTGGCCCCAGACTTTGTTGC 643

 199 yAspValLeuHisAlaHisLeuAspIleTrpLys..............  210
     ||||:::|||      |||||||:::|||   :::
 644 GGACATCTTATGGGAACATCTAGAGATCCTGCAGAAAGAGGATGTGAAAC 693

 211 ....................SerAlaAlaTrpMetLysGluArg...Thr  219
                         :::   |||:::   :::::|||   |||
 694 CATATCAAGTTAATGGAGTCAACCCAGCCTATCCAGAATCCCGCTATACC 743

 220 SerProGlyAlaIleHisTyr.................CysAlaSerTh  230
     |||       |||   |||                  ||||::
 744 TCGGATTACTTCATTAGCTATGGTATTGAGCATGCCCAGTCTGTTCCACC 793

 230 rSerGlu.......................GluSerTrp..........  235
      ||||||                     |||||||:::
 794 ATCGGAGTTCTCAGAGCCCAGCTTCATCACAGAGTCCTATCAGACGCTCC 843

 235 ..................................................  235

 844 ATCCCATCAGCTCGGAAGAGCTCCTCTCCCTCAAGTATGAGAATGACTAC 893

 236 ..............................ThrAspSer...........  238
                                   |||||||:::
 894 CCCTCGGTCATTCTCCGAGACCCTCTCCAGACAGACACCTTGCAGAATGA 943

 239 ...................GluVal.........AspSerSerCysSerG  246
                        ||||||         ||:::   |||   |
 944 CTACTTTGCTATCAAACAAGAAGTCGTCACCCCAGACAACATGTGCAIGG 993

 246 lyGln..                                           247
     ||:::
 994 GGAGGACCAGTCGTGGTAAACTCGGGGGCCAGGACTCTTTTGAAAGCATA 1043

 247                                                   247

1044 GAGAGCTACGATAGTTGTGATGGGCTCACCCAGTCCTGGAGCAGCCAGTC 1093

 247 ..................................................  247

1094 ATCTTTCAACAGCCTGCAGCGTGTTCCCTCCTATGACAGCTTCGACTCAG 1143

 247 ..................................................  247

1144 AGGACTATCCGGCTGCCCTGCCCAACCACAAGCCCAAGGGCACCTTCAAG 1193

 247 ..................................................  247

1194 GACIATGTGCGGGACGGTGCTGACCTCAATAAGGACAAGCCTGTCATTGC 1243

 248 ..............................ProIleHisLeuTrpG  253
                                  |||||||:::|||||||
1244 TGCTGCTGCCCTAGCTGGCTACACAGGCAGTGGACCAATCCAGCTATGGC 1293

 253 luPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl  269
     ||||||||   |||   ||||||      |||          ||||||
1294 AGTTTCTTCTGCAA   TTACTCACTGATAAATCCTGTCAGTCTTTTATC 1340

 270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl  286
     |||   :::   :::          ||||||:::::::::          ::
1341 AGCTGGACAGGA...GATGGCTGGGAATTCAAACTTTCTGACCCAGATGA 1387

 286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA  303
     :|||||||||   ||||||   |||||||||:::|||   |||||||||:
1388 GGTGGCCAGGAGATGGGGAAAGAGGAAAAACAAACCTAAGATGAATTATC 1437

 303 spLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArg  319
     ::|||||||||||:::::|||   ||||||   |||:::|||||||:::
1438 AGAAACTGAGCCGTGCCCTACGCTACTATTACGACAAAAACATCATCCAC 1487

 320 LysProAspIleSerGlnArgLeuValTyrGlnPheVal  332
     |||         :::::|||   ||||||::||||||
```

```
     1488 AAG   ACAGCGGGCAAACGCTACGTGTACCGCTTTGTG 1523

seq number /seq1 type nuc 5' 1 3' 1920 SA COMB seq:US-08-746-789A-1

seq_documentation_block:
; Sequence 1, Application US/08746789A
; Patent No. 5789200
;  GENERAL INFORMATION:
;    APPLICANT:  Ismail Kola, Martin J. Tymms, Christine DeBouck
;    TITLE OF INVENTION:  A No. 5789200el Human ETS Family Member, ELF3
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road, P.O. Box 1539
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 5.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM 486
;      OPERATING SYSTEM:  WINDOWS FOR WORKGROUPS
;      SOFTWARE:  MICROSOFT WORD
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/746,789A
;      FILING DATE:  No. 5789200ember 15, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William T. Han
;      REGISTRATION NUMBER:  34,844
;      REFERENCE/DOCKET NUMBER:  AIG 50024
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610 270 5219
;      TELEFAX:  610 270 4026
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1920
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    ANTI-SENSE:  No
US-08-746-789A-1


alignment_scores:
            Quality:  250.00              Length:    429
              Ratio:  1.474                 Gaps:     17
 Percent Similarity:  42.424   Percent Identity:  22.844

alignment_block:
 US-09-126-945B-2 x US-08-746-789A-1    ..

 Align_Seq 1/1  to: US-08-746-789A-1  from: 1  to: 1920

       31 AlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPr  47
          |||||||||      ::::::          |||:::|||          ||
       16 GCAGCGGCCAGATACCTCAGCGCTACCTGGGGGAACTGGATTTCTCTCCC  65

       47 oPro.....................................................  48
          ||||
       66 GCCTGCCCGCCTCCCTGCCACAGCCCCACTCCGCCACCCCCTACGCTCCA  115

       49 .....AlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPhe  83
               ||||||   |||      ::::||   ::::::   ::||||||||
      116 TGGCTGCAACCTGTGAG......ATTAGCAACATTTTTAGCAACTACTTC  159

       84 AspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAl  80
          :::   :::|||   ::::::::::||      |||::::::|||   ||
      160 AGTGCGATGTACAGCTCGGAGGACTCCAGCTTGGCCTCTGTTCCCCCTGC  209
```

```
 80 aSerSerArgGluGluProProGluGluProGlnGlnCysProValIleA 97
    |
210 t.................................................. 210

 97 spSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThr 113
           |||      ||||::  |||||||||         ||||||
211 ........GCCACCTTTGGGGCGATGACTTGGTA.........CTGACC 243

114 LeuGln......GluHisSerLeuGluGlnValGln...SerMetValVa 127
    ||||::      :::   |||||||||   ::::::  :::     ::
244 CTGAGCAACCCCCAGATGTCATTGGAGGGTACAGAGAAGGCTAGCTGGTT 293

127 lGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleI 144
    :||||||
294 GGGGGAA.............................................. 300

144 hrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeu 160
       :::|||:::   ||||||   :::::::|||      ||||::
301 .....CAGCCCCAGTTCTGGTCGAAGACGCAGGTTCTGGACTGGATCAGC 345

161 Trp......ThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPh 175
    :::         ::::::::::|||          :::      :::
346 TACCAAGTGGAGAAGAACAAGTACGACGCAAGCGCCATTGACTTCTCACG 395

175 sGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGlnGluGlnPheA 192
       ::::::   |||      ||||||      :::   ||||::::::|
396 ATGTGACATGGATGGGGCCACCCTCTGCAATTGTGCCCTTCAGGAGCTGC 445

192 rq...GlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu... 206
    ||         :::||||||   ||||||   ||||||||||:::|||
446 GTCTGGTCTTTGGGCCTCTG...GGGGACCAACTCCATGCCCAGCTGCGA 492

207 AspIleTrpLysSerAla.............AlaTrpMet........... 215
    |||:::   :::|||:::            :::|||:::
493 GACCTCACTTCCAGCTCTTCTGATGAGCTCAGTTGGATCATTGAGCTGCT 542

216 .......................LysGluArgThrSerProGlyAlaIleH 225
                           :::|||      :::||||||   :::
543 GGAGAAGGATGGCATGGCCTTCCAGGAGGCCCTAGACCCAGGGCCCTTTG 592

225 isTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluValAsp 241
       :::      :::::::|||          |||::::::
593 ACCAGGGCAGCCCCTTTGCCCAGGAGCTGCTGGACGACGGTCAGCAAGCC 642

242 Ser................SerCysSer...................... 245
    |||               |||||:::
643 AGCCCCTACCACCCCGGCAGCTGTGGCCCAGGAGCCCCCTCCCCTGGCAG 692

245 .................................................. 245

693 CTCTGACGTCTCCACCGCAGGGACTGGTGCTTCTCGGAGCCCCCACTCCT 742

245 .................................................. 245

743 CAGACTCCCGGTGGAAGTGACGTGGACCTGGATCCCACTGATGGCAAGCTC 792

245 .................................................. 245

793 TTCCCCAGCGATGGTTTTCGTGACTGCAAGAAGGGGGATCCCAAGCACGG 842

246 ............GlyGlnPro.......................... 248
               |||::|||
843 GAAGCGGAAACGAGGCCGGCCCCGAAAGCTGACCAAAGAGTACTGGGACT 892

249 ..................................IleHisLeuTrp 252
                                     |||||||||
893 GTCTCGAGGGCAAGAAGAGCAACCACGGCCCAGAGGCACCCACCTGTGG 942

253 GluPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheI 269
    :::|||:::::::::::::|||:::   |||:::          :::::
943 GAGTTCATCCGGGACATCCTCATCCACCCGGAGCTCAACGAGGGCTCAT 992

269 rArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaC 285
```

```
       :::::|||   |||::::::::::|||:::|||||||:::      |||
    993 GAAGTGGGAGAATCGGCATGAAGGCGTCTTCAAGTTCCTGCGCTCCGAGG 1042

    286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
          |||||:::|||||||||   :::|||:::::::      |||:::|||
   1043 CTGTGGCCCAACTATGGGGGCAAAAGAAAAAGAACAGCAACATGACCTAC 1092

    303 AspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 319
        :::|||||||||||||:::::|||   ||||||||||:   |||:::::
   1093 GAGAAGCTGAGCCGGGGCCATGAGCTACTACTACAAACGGGAGATCCTGGA 1142

    319 qLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
        ::::   |||   ::::::|||||||||||||:::|||
   1143 ACGGCTGGAT...GGCCGGCGACTCGTCTACAAGTTT 1176

seq_name: Seq1, Sep-18-01 1:01 'A_C' MIC seq: US-08-368-281-1

seq_documentation_block:
; Sequence 1, Application US/08368281
; Patent No. 5721113
;  GENERAL INFORMATION:
;     APPLICANT:  Libermann, Towia A
;     APPLICANT:  Oettgen, Joerg P
;     APPLICANT:  Kunsch, Charles A
;     TITLE OF INVENTION:  NERF Genes
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporation
;       STREET:  709 Swedeland Road
;       CITY:  King of Prussia
;       STATE:  Pennsylvania
;       COUNTRY:  USA
;       ZIP:  19406
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/368,281
;       FILING DATE:
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Baumeister, Kirk
;       REGISTRATION NUMBER:  33,833
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  610-270-5096
;       TELEFAX:  610-270-5090
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2975 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-368-281-1


alignment_scores:
            Quality:  242.00              Length:    261
              Ratio:  1.847                 Gaps:      8
 Percent Similarity:  50.192   Percent Identity: 25.287

alignment_block:
 US-09-126-945B-2 x US-08-368-281-1  ..

 Align seq 1/1  to: US-08-368-281-1  from: 1  to: 2975

      99 GlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGl 115
         ::::::|||:::::::::|||||||||:::     ::::::      ::
     140 GAGGGACCCACGAACCAGCTGGATCTGCTCATCCGGGGCCGTGGAAGCATC 189

     115 nGluHisSer.....................LeuGlnGlnValG 123
```

```
      :   ||||||                                :::|||   ::::
  190 AGTTCACAGCAGTAATGCACACTGTACAGATAAGACAATTGAAGCTGCTG 239

  123 lnSerMetValValGlyGln.........ValLeuLysAsp......... 148
      ::::::::::      |||           ||||:|||
  240 AAGCCCTGCTTCATATGGAATCTCCTACCTGCTTGAGGGATTCAAGAAGT 289

  134 .....IleGluThrAl 147
                                                ||||||::
  290 CCTGAATTCATCCATGCTGCCTATGAGGCCAGATGTCATTACAGAAACTGT 339

  137 aCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerA 184
      :   ::::::   :::      ::::::||||||||   ||||||
  340 AGTGGAGGTGTCAACTGAAGAGTCTGAACCCATGGATACCTCTCCT.... 385

  154 snValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProPro 178
                                                :::|||
  386 .....ATTCCAACA 394

  171 MetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSe 187
          :::   :::   :::      ||:::          :::::
  395 TCACCAGATAGCCATGAACCAATGAAAAAGAAAAAA.....GTTGG 437

  187 rGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuHisA 204
      :::::::      ||||:::|||||:::::::::
  438 CCGTAAACCAAAGACCCAGCAATCACCAATTTCCAATGGG.......... 475

  204 laHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSer 220
                      :::                 :::|||:::::
  476 ..............TCTCCTGAGTTAGGTATAAAGAAGAAACCAAGA 508

  221 ProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAs 237
        |||
  509 GAAGGA..................................... 514

  237 pSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnP 254
                      ::::|:::      :::|||||||||:|
  515 ....................AAAGGAAACACAACCTATTTGTGGGAGT 542

  254 heLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArg 270
      |||||   ::::||||||          ::::::      ||||:|||:::
  543 TTCTTTTAGATCTACTTCAAGATAAAAATACTTGTCCCAGGTATATTAAA 592

  271 TrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVa 287
      |||   ::::::||||||||||||||||||||:::   ||||||     ||
  593 TGGACTCAGAGAGAAAAAGGCATATTCAAGCTGGTGGATTCAAAGGCTGT 642

  287 lAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspL 304
      |::::::||||||||   :::||||||:::||   ||||||||||:::
  643 CTCTAAGCTTTGGGGAAAGCATAAGAACAAACCAGACATGAACTATGAAA 692

  304 ysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLys 320
      :::::::|||:::::::||   ||||||:::::::|||||||:::   |||
  693 CCATGGGACGAGCTTTGAGATACTACTACCAAAGGGGAATTCTTGCAAAG 742

  321 ProAspIleSerGlnArgLeuValTyrGlnPhe 331
        :::   :::|||||||||||||||||||||
  743 GTTGAA...GGACAGAGGCTTGTATATCAGTTC 772

seq_summary_for_US-09-126-945B-2 x US-09-009-913-338

seq_documentation_block:
; Sequence 338, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;     APPLICANT:  AxyS Pharmaceuticals, Inc.
;     TITLE OF INVENTION:  Asthma Related Genes
;     NUMBER OF SEQUENCES:  339
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Bozicevic & Reed, LLP
;       STREET:  285 Hamilton Ave, Suite 200
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94301
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/009,913
;       FILING DATE:  21-JAN-1998
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sherwood, Pamela J
;       REGISTRATION NUMBER:  36,677
;       REFERENCE/DOCKET NUMBER:  SEQ-4P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-327-3231
;       TELEFAX:  650-327-3231
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  338:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  848 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  Coding Sequence
;       LOCATION:  1...848
;       OTHER INFORMATION:
US-09-009-913-338


alignment_scores:
            Quality:  239.00              Length:    237
              Ratio:  1.897                 Gaps:      8
 Percent Similarity:  53.165   Percent Identity: 27.004

alignment_block:
 US-09-126-945B-2 x US-09-009-913-338  ..

 Align seq 1/1  to: US-09-009-913-338  from: 1  to: 848

  147 ProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGl 163
      |||:::   ||:::       :::|||   :::||||||
  139 CCTCAGTACTGGACCAAATACCAGGTGTGGGAATGGCTGCAGCACCTCCT 188

  163 uHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu......L 178
                  ::::|||                |||||||||      :
  189 GGACACCAACCAGCTAGACGCTACGTGCATCCCTTTCCAGGAGTTCGACA 238

  178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArg 194
      .....|||.....:|||||:::||||||   :::::|||   :::
  239 TTAGCGGAGAACACCTGTGCAGCATGAGTCTGCAGGAGTTCACGAGGGCA 288

  195 SerProLeuGlyGlyAspValLeuHisAlaHisLeu..............  206
      :::       ::::|||:::::::::|||||:::::::|||
  289 GCAGGCTCAGCTGGGCAGCTGCTCTACAGCAACCTACACCATCTCAAGTG 338

  207 ..          ..........  AspTrpTrpLysSerAla          AlaT 214
                              |||:::::::::||||||
  339 GAACGGCCAATGCAGCAGTGACCTTTTCCAGTCCGCACACAATGTCATTG 388

  214 rpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThr 230
            |||:::|||:::|||:::   ::::::         :::
  389 TCAAGACTGAACAAACCGATCCTTCCATCATGAAC.........ACATGG 429

  231 SerGluGluSerTrpThr.............AspSerGluValAsp.. 241
      :::||||||:::::                |||   ||||||
  430 AAAGAAGAAAACTATCTCTATGATCCCAGCTATGGTAGCACAGTAGATCT 479
```

```
     242 ..............SerSerCysSerGlyGln..................... 247
                        :::   :::   :::|||
     480 GTTGGACAGTAAGACTTTCTGCCGGGCTCAGATCTCCATGACAACCTCCA 529

     247 ..................................................... 247

     530 GTCACCTTCCAGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCAC 579

     248 ProIle.....................................HisLeuIl 262
         |||:::                                    ||||||||
     580 CCTGTAAAGTCCCACACCAAAAAGCACAACCCAAGAGGCACTCACTTATC 629

     252 pGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheI 269
         |:::|||:::::::::::::|||||:::|||   :::          :::|
     630 GGAGTTCATCCGAGACATTCTCTTGAGCCCAGACAAGAACCCAGGGCTGA 679

     269 IeArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAla 285
         |||:::|||   :::::::::::::|||||||||::::::      |||
     680 TCAAATGGGAAGACCGTTCGGAAGGCATCTTCAGGTTCCTGAAGTCAGAA 729

     286 GlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTy 302
         |||||||:::|||||||||   :::|||||||:::   :::|||:::||
     730 GCTGTGGCTCAGCTGTGGGGGAAAAAGAAAAATAACAGTAGCATGACATA 779

     302 rAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleA 319
         |:::|||||||||||||:::::|||   ||||||||||:::   |||:::
     780 CGAGAAGCTCAGCCGGGCTATGAGATATTACTACAAACGAGAAATCCTGC 829

     319 rgLysProAsp 322
         :::::   |||
     830 AACCTGTGGAT 840

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-875-944B-1

seq_documentation_block:
; Sequence 1, Application US/08875944B
; Patent No. 6096542
;  GENERAL INFORMATION:
;    APPLICANT:  FUJINAGA, Kei
;    APPLICANT:  YOSHIDA, Koichi
;    APPLICANT:  HIGASHINO, Fumihiro
;    TITLE OF INVENTION:  CANCER CONTROL
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK, P.L.L.C.
;      STREET:  624 Ninth Street N.W., Ste. 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/875,944B
;      FILING DATE:  07 AUG 1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 07 020173
;      FILING DATE:  08 FEB 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP96/00016
;      FILING DATE:  09-JAN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  FUJINAGA 1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 628-5197
;      TELEFAX:  (202) 737-3528
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2064 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1386
US-08-875-944B-1


alignment_scores:
            Quality:  230.50              Length:     357
              Ratio:  1.332                 Gaps:      21
 Percent Similarity:  48.459   Percent Identity: 28.571

alignment_block:
 US-09-126-945B-2 x US-08-875-944B-1   ..

 Align seq 1/1  to: US-08-875-944B-1  from: 1  to: 2064

      44 SerProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLe 60
         |||   :::|||||          :::::|||   ::::::::|||
     265 AGCAGGAAGCCGCCACTCCCCTACCACCATGGCGAGCAGTGCCTTTACTC 414

      60 uSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysA 77
          |||   :::|||         |||::::::   :::      :::   |
     315 CAGTGCCTATGACCCC.....CCCAGACAAATCGCCATCAAGTCCCCTG 458

      77 laProGlyAla......SerSerArgGluGluProProGluGlnProGlu 91
         ||||||||||||      :::|||     :::|||      :::   |||
     359 CCCCTGGTGCCCTTGGACAGTCGCCCCTACAGCCCTTTCCCCGGGCAGAG 408

      92 GlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuVa 108
         |||          :::   |||         :::|||:::   :::
     409 CAACGGAATTTCCTGAGATCC.....TCTGGCACCTCCCAGCCCCA 449

     108 lProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerM 125
          ||||||         |||   |||||||||
     450 CCCTGGCCATGGGTACCTCGGGCAACATAGC..................... 480

     125 etValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeu 141

     480 ..................................................... 480

     142 AsnIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTr 158
         ::::::      :::|||:::|||         |||
     481 TCCGTCTTCCAGCAGCCCCTGGACATTTGCCACTCC.............. 516

     158 pLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaP 175
                ::|||:::::::                |||:::      :
     517 .......TTCACATCTCAGGGAGGGGCCCGGCAACCCCTCCCAGCCCCCT 559

     175 heGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPhe 191
         :::|||::      ::::::                  ::::::::|||
     560 ACCAACACCAGCTGTCGCAGCCCTGCCCACCCTATCCCCAGCAGAGCTTT 609

     192 ArgGln......ArgSerProLeu.......................... 197
         :::|||      :::::||||||
     610 AAGCAAGAATACCATGATCCCCTGTATGAACAGGCGGGCCAGCCAGCGT 659

     198 .......GlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerA 212
              ||||||      ::::::::::|||         :::   :::|
     660 GGACCAGGCTGGG......GTCAATGGGCACAGG......TACCCAGGGG 697

     212 la.AlaTrpMetLysGluArgThr........................... 219
         || ::|||   ::::::::::::
     698 CGGGGGTGGTGATCAAACAGGAACAGACGGACTCGCCTACGACTCAGAT 747

     220 .SerProGlyAlaIleHistyrCysAlaSerThrSerGluGluSer.Trp 235
          ||||||||||||   |||   |||:::||||||::::::   ||| |||
```

```
     748 GTCACCGGGTGCG...CATCAATGTACCTCCACACAGAGGGCTTCTCTGG 794

     236 ThrAspSerGlu..................................... 239
         :::   |||:::
     795 GCCCTCTCCAGGTGACGGGGCCATGGGCTATGGCTATGAGAAAGCCTCTGC 844

     240 .......Val.AspSerSerCys.......................... 244
                 |||:::   |||
     845 GACCATTCCCAGATGATGTCTGCGTTGTCCCTGAGAAATTTGAAGGAGAC 894

     245 ............................SerGlyGlnPro........ 248
                                    :::|||   |||
     895 ATCAAGCAGCAAGGGGTCGGTGCATTTCGAGAGGGGCCGCCCTACCAGCG 944

     249 ..........IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysP 262
                  :::::|||||||||||||||||     ||||||      |
     945 CCGGGGTGCCCTGCAGCTGTGGCAATTTCTGGTGGCCTTGCTGGATGACC 994

     262 roHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
         ||   :::   :::::|||||    |||   :::       :::|||:::
     995 CA...ACAAATGCCCATTTCATTGCCTGGACGGGC......CGGGGAATG 1035

     279 ...PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleAr 294
            |||||:::   :::       :::|||||||||||||||||||||::
    1036 GACTTCAAGCTCATTGAGCCTGAGGAGGTCGCCAGGCTCTGGGGCATCCA 1085

     294 gLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgG 311
         :|||||||:|||||||||||||||||||||||||||||||||||||:::|||
    1086 GAAGAACCGGCCAGCCATGAATTACGACAAGCTGAGCCGCTCGCTCCGAT 1135

     311 lnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeu 327
           ||||||:::|||||||:::::|||        :::::|||
    1136 ACTATTATGAGAAAGGCATCATCCAGAAGGTG   GCTCGTGAGCGTTAC 1182

     328 ValTyrGlnPheVal 332
         ||||||:::||||||
    1183 GTGTACAAGTTTGTG 1197

seq_name: /cgn2_9/ptodata/1/rna/68_COMB.seq:US-09-116-049-3

seq_documentation_block:
; Sequence 3, Application US/09116049A
; Patent No. 6248351
; GENERAL INFORMATION.
;  APPLICANT: Hung, Men-Chie
;  TITLE OF INVENTION: HUMAN PEA3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS
;  FILE REFERENCE: UTSC:582
;  CURRENT APPLICATION NUMBER: US/09/116,049A
;  CURRENT FILING DATE: 1998-07-15
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2064
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-116-049-3


alignment_scores:
            Quality:  230.50              Length:    357
              Ratio:  1.332                 Gaps:     21
 Percent Similarity:  48.459   Percent Identity: 28.571

alignment_block:
 US-09-126-945B-2 x US-09-116-049-3

 Align seq 1/1  to: US-09-116-049-3  from: 1  to: 2064

       44 SerProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLe 60
          |||   :::|||||||          :::::::|||   ::::::::|||
      265 AGCAGGAAGCCGCCACTCCCCTACCACCATGGCGAGCAGTGCCTTTACTC 314

       60 uSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysA 77

     |||   :::|||          |||:::::::   :::       :::   |
315 CAGTGCCTATGACCCC.....CCCAGACAAATCGCCATCAAGTCCCCTG 458

 77 laProGlyAla......SerSerArgGluGluProProGlnGluProGlu 91
    ||||||||||||      :::|||      :::||||      :::   |||
359 CCCCTGGTGCCCTTGGACAGTCGCCCCTACAGCCCTTTCCCCGGGCAGAG 408

 92 GlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuVa 108
    |||          :::   |||          :::|||:::   :::
409 CAACGGAATTTCCTGAGATCC.........TCTGGCACCTCCCAGCCCCA 449

108 lProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerM 125
     ||||||       |||   |||||||||
450 CCCTGGCCATGGGTACCTCGGGGAACATAGC..................... 480

125 etValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeu 141

480 ..................................................... 480

142 AsnIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTr 158
    ::::::      :::|||:::|||          |||
481 TCCGTCTTCCAGCAGCCCCTGGACATTTGCCACTCC.............. 516

158 pLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaP 175
          :::|||::::::                 ||||::      :
517 .......TTCACATCTCAGGGAGGGGGCCGGGAACCCCTCCCAGCCCCCT 559

175 heGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPhe 191
    ::|||:::      ::::::                          ::::::::|||
560 ACCAACACCAGCTGTCGGAGCCCTGCCCACCCTATCCCCAGCAGAGCTTT 609

192 ArgGln .     ArgSerProLeu  .  .  .  .  .  .  .  . 197
    ::|||       ::::::||||||
610 AAGCAAGAATACCATGATCCCCTGTATGAACAGGCGGGCCAGCCAGCCGT 659

198 .......GlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerA 212
          ||||||      :::::::::|||        :::   :::|
660 GGACCAGGGTGGG      GTCAATGGGCACAGG      TACCCAGGGG 697

212 la.AlaTrpMetLysGluArgThr.................... 219
    || :::|||.  ::::::|||:::
698 CGGGGTGGTGATCAAACAGGAACAGACGGACTTCGCCTACGACTCAGAT 747

220 .SerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSer.Trp 235
     ||||||||||||   |||   |||:::||||||:::::   ||| |||
748 GTCACCGGGTGCG   CATCAATGTACCTCCACACAGAGGGCTTCTCTGG 794

236 ThrAspSerGlu                ........................ 239
    :::   |||:::
795 GCCCTCTCCAGGTGACGGGGCCATGGGCTATGGCTATGAGAAAGCCTCTGC 844

240 .......Val.AspSerSerCys.......................... 244
            |||:::   |||
845 GACCATTCCCAGATGATGTCTGCGTTGTCCCTGAGAAATTTGAAGGAGAC 894

245 ............................SerGlyGlnPro........ 248
                               :::|||   |||
895 ATCAAGCAGCAAGGGGTCGGTGCATTTCGAGAGGGGCCGCCCTACCAGCG 944

249 ..........IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysP 262
            :::::|||||||||||||||     ||||||      |
945 CCGGGGTGCCCTGCAGCTGTGGCAATTTCTGGTGGCCTTGCTGGATGACC 994

262 roHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
    ||   :::   :::::|||||    |||   :::       :::|||:::
995 CA...ACAAATGCCCATTTCATTGCCTGGACGGGC......CGGGGAATG 1035

279 ...PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleAr 294
       |||||:::   :::       :::|||||||||||||||||||||::
1036 GAGTTCAAGCTCATTGAGCCTGAGGAGGTCGCCAGGCTCTGGGGCATCCA 1085

294 gLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgG 311
    :|||||||||||||||||||||||||||||||||||||||||||||:::|||
```

```
    1086 GAAGAACCGGCCAGCCATGAATTACGACAAGCTGACGCGCTCGCTCCGAT 1135

     411 InTyrTyrLysLysGlyIleIleArgLysProAspIleSerGluArgLeu 427
         ||||||:::|||||||||:::::|||         :::::|||
    1146 ACTATTATCAGAAAGGCATCATCCAGAAGGTG...GCTGGTGAGCGTTAC 1192

     428 ValTyrGlnPheVal 442
         ||||||:::||||||
    1193 GTGTACAAGTTTGTG 1207

Seq name: /equn2_6/prodata/1/ina/5B_COMB_seq-US-08-780-835B-1

Seq documentation block:
; Sequence 1, Application US/08780835B
; Patent No. 5922688
;  GENERAL INFORMATION:
;    APPLICANT: Hung, Mien-Chie
;    APPLICANT: Xing, Xianqming
;    TITLE OF INVENTION: PEA3 is a Tumor Suppressor
;    NUMBER OF SEQUENCES: 9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: ARNOLD, WHITE AND DURKEE
;      STREET: P.O. Box 4433
;      CITY: Houston
;      STATE: Texas
;      COUNTRY: USA
;      ZIP: 77210-4433
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC DOS/MS DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/780,835B
;      FILING DATE: 10-JAN-1997
;      CLASSIFICATION: 514
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Wilson, Mark B.
;      REGISTRATION NUMBER: 37,259
;      REFERENCE/DOCKET NUMBER: UTSC500
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (512) 418-3000
;      TELEFAX: (512) 474-7577
;  INFORMATION FOR SEQ ID NO: 1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 2410 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
US-08-780-835B-1


Alignment scores:
            Quality:  229.50              Length:    491
              Ratio:    1.158               Gaps:     26
 Percent Similarity:   40.530   Percent Identity: 24.440

Alignment block:
 US-09-126-945B-2 x US-08-780-835B-1  ..

 Align seq 1/1  to: US-08-780-835B-1  from: 1  to: 2410

      12 SerProSerHisLeuLeuLeuProProAspThrValSerArgThr..... 26
         ||||||:::|||       ||||||   :::   :::::|||
     243 TCCCCCGGGCATGCTGAGCATCCCCCCCGGCTCCGGCACAGACGCCGG 292

      27 .........................GlyLeuGluLysAlaAlaAlaGlyA 35
                                  |||         :::|||:::|||
     293 ACCTCAGGTCTCTGCCTCCGCGCGGGGGCCGGCCCTGTGGCCGGAGGGA 342

      35 laValGlyLeuGluArgArg............................. 41
         ::      :::|||||||||
     443 GCGGCCGGATGGAGCGGAGGATGAAAGGCGGATACTTGGACCAGCGAGTG 392

      41 ..................................................... 41

     393 CCCTACACCTTCTGCAGCAAATCTCCCGGAAATGGGAGCTTGGGCGAAGC 442

      42 ..........................AspTrpSerProSerProProA 49
                                  |||   :::     ||||||
     443 GCTGATGGTCCCGCAGGGAAAGCTCATGGACCCGGGCTCCCTGCCGCCTT 492

      49 laThrProGlu.........GluGlyLeuSerAlaPheTyrLeuSerTyr 62
         ::      |||         |||   ||||||   |||
     493 CCGACTCAGAAGATCTCTTCCAGGACCTCAGTCACTTC............ 530

      63 PheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla..... 77
                             ::::::::::|||   |||:::|||
     531 ..................CAAGAGACGTGGCTCGCAGAAGCTCAGGT 559

      77 ..................................................... 77

     560 ACCGGACAGTGATGAGCAGTTTGTTCCTGATTTCCATTCAGAAAACTTAG 609

      77 ..................................................... 77

     610 CTTTCCATAGCCCCACCACCAGGATCAAGAAGCAACCCCAGAGTCCCCGC 659

      78 ......ProGlyAlaSerSerArgGluGluProPro.........GluGl 89
              ||:::   |||       :::::||||||          :::::
     660 ACAGACCCCGCCCTGTCCTGCAGCAGGAAGCCACCACTCCCCTACCACCA 709

      89 uProGluGluCys..............ProValIleAspSerGlnAlaP 101
         :  ||||||||                :::|||  |||  ||||
     710 TGGAGAGCAGTGCCTTTACTCCAGACAAATCGCCATCAAGTCCCCCGCTC 759

     101 roAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHis 117
         ||:::               :::||||||         |||:::
     760 CCGGT..............GCCCCTGGACAGTCGCCCCTGCAGCCCTTT 794

     118 Ser...LeuGluGlnValGluSerMetValValGlyGlnValLeuLysAs 133
         |||      ||||||   |||||:::::::   ::::::      :::::
     795 TCCAGGGCAGAACAGCAGCAGAGCCTCCTGAGAGCCTCCAGCTCTTCCCA 844

     134 pIleGluThrAlaCysLysLeuLeu.............AsnIleThrAlaA 146
         :  :::   :::          |||                ::::::     :
     845 GTCCCACCCTGGCCACGGGTACCTTGGTGAGCACAGCTCCGTCTTCCAGC 894

     146 spProMetAspTrpSerProSerAsnValGluLysTrpLeuLeuTrpThr 162
         ::|||:::|||
     895 AGCCCGTGGACATGTGC.............................. 911

     163 GluHisGlnTyrArgLeuPro.................ProMetGlyLy 173
            |||::::::::   |||               ||||::
     912 ...CACTCCTTCACATCTCCTCAGGGAGGGGGCCGGGAACCTCTCCCAGC 958

     174 sAlaPheGluGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluG 190
         :::|||:::      ::::::                ::::::     :
     959 CCCCTATCAACACCAACTGTCGGAGCCCTGCCCACCCTACCCCCAGCAGA 1008

     190 lnPheArgGlnArg..................................... 194
         ::|||:::|||:::
    1009 ACTTCAAGCAGGAGTACCATGACCCCCTGTACGAACAGGCTGGCCAGCCC 1058

     195 ...SerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLy 210
            |||      ||||||       |||                :::
    1059 GCTTCAAGCCAGGGTGGGGTCAGTGGGCACAGG.....TACCC 1096

     210 sSerAla.AlaTrpMetLysGluArg.........IhrSerProGlyAla 223
         ::|||  :::|||   :::::::::|          |||||||||   :::
    1097 AGGGAGCGGGAGTGGTGATCAAACAGGAGCGCACAGACTTCGCCTACGACT 1146

     224 ................IleHisTyrCysAlaSerThrSerGluGluSerI 235
                       :::|||   |||:::||||||:::::   |||:|
    1147 CAGATGTCCCTGGATGTGCATCAATGTACCTCCACCCAGAGGGCTTCTCT 1196
```

```
     235 rpThrAspSerGlu..................................... 239
         |||||   |||:::
    1197 GGACGCTCTCCAGGTGATGGAGTGATGGGTTATGGCTATGAAAAATCCCT 1246

     240 .........Val.AspSerSerCys.......................... 244
                  |||:::   |||
    1247 TCGACCATTCCCAGATGATCTCTGCATTGTCCCTAAAAAATTTGAAGGAG 1296

     245 ................................SerGlyGlnPro...... 248
                                        :::|||   |||
    1297 ACATCAAGCAGGAAGGGATTGGAGCTTTCCGGGAGGGGCCACCCTACCAG 1346

     249 ............IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLy 261
                     :::::|||||||||||||||||      ||||||
    1347 CGCCGGGGTGCCTTACAACTGTGCCAGTTTCTGGTGGCCCTGCTGCATGA 1396

     261 sProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyI 278
          |||   :::   ::::::|||||   |||   :::       ::::|||
    1397 CCCA...ACAAATGCTCATTTCATTGCTTGGACAGGC......CGGGGAA 1437

     278 le...PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIle 293
         ::   |||||:::   :::       :::|||||||||||||||||||||
    1438 TGGAGTTTAAACTAATTCAACCTGAAGAGGTTGCCAGGCTCTGGGGTATC 1487

     294 ArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleAr 310
         :::|||||||||||||||||||||||||||||||||||||||||||:::||
    1488 CAGAAGAACCGGCCAGCCATGAATTATGACAAGCTGAGCCGCTCGCTGCG 1537

     310 qGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGluArgL 327
         |   ||||||:::|||||||||:::::|||          :::::|||
    1538 ATACTATTATGAGAAAGGCATCATGCAGAAGGTG...GCTGGCGAACGCT 1584

     327 euValTyrGlnPheVal 332
         |||||:::|||||
    1585 ACGTGTACAAGTTTGTG 1601

Seq_id#1: 2002 bp; Type=DNA; Strand=Plus/Plus; GB_COMB seq=US-09-303-268-1

seq_documentation_block:
; Sequence 1, Application US/09303268
; Patent No. 6172212
;   GENERAL INFORMATION:
;        APPLICANT: Hung, Mien-Chie
;                   Xing, Xiangming
;        TITLE OF INVENTION: PEA3 is a Tumor Suppressor
;        NUMBER OF SEQUENCES: 9
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: ARNOLD, WHITE AND DURKEE
;             STREET: P.O. Box 4433
;             CITY: Houston
;             STATE: Texas
;             COUNTRY: USA
;             ZIP: 77210-4433
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/303,269
;             FILING DATE: 30-Apr-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/780,835
;             FILING DATE: 10-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Wilson, Mark B.
;             REGISTRATION NUMBER: 37,259
;             REFERENCE/DOCKET NUMBER: UTSC500
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (512) 418-3000
;             TELEFAX: (512) 474-7577
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2410 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-303-268-1


alignment_scores:
            Quality:  229.50              Length:    491
              Ratio:  1.153                 Gaps:     26
 Percent Similarity:  46.530   Percent Identity: 24.440

alignment_block:
 US-09-126-945B-2 x US-09-303-268-1   ..

 Align seq 1/1  to: US-09-303-268-1  from: 1  to: 2410

      12 SerProSerHisLeuLeuLeuProProAspThrValSerArgThr..... 26
         |||||||:|||       ||||||   :::   :::::|||
     243 TCCCCCGGGCATGCTGAGCATCCCCCCGCGGCTCCGGCACAGACGCCCGG 292

      27 .........................GlyLeuGluLysAlaAlaAlaGlyA 45
                                  |||         ::||||:::||||
     293 ACCTCAGGTCTCTGCCTCCGCGCGGGGGCCCGGCCCTGTGGCCGGAGGGA 342

      35 laValGlyLeuGluArgArg.............................. 41
         ::      :::|||:|||||
     343 GCGGCCGGATGGAGCGGAGGATGAAAGGCGGATACTTGGACCAGCGAGTG 392

      41 ................................................... 41

     393 CCCTACACCTTCTGCAGCAAATCTCCCGGAAATGGGAGCTTGGGCGAAGC 442

      42 ............................AspTrpSerProSerProProA 49
                                     |||   :::     |||||||
     443 GCTGATGGTCCCGCAGGGAAAGCTCATGGACCCGGGCTCCCTGCCGCCTT 492

      49 laThrProGla          GlnGlyLeuSerAlaPheTyrLeuSerTyr 62
         ::     |||          |||   ||||||   |||
     493 CCGACTCAGAAGAICTCTTCCAGGACCICAGTCACTTC.............. 530

      63 PheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla..... 77
                              ::::::::::||   |||:::|||
     531 .....................CAAGAGACGTGGCTCGCAGAAGCTCAGGT 559

      77 ................................................... 77

     560 ACCGGACAGTGATGAGCAGTTTGTTCCTGATTTCCATTCAGAAAACTTAG 609

      77 ................................................... 77

     610 CTTTCCATAGCCCCACCACCAGGATCAAGAAGGAACCCCAGAGTCCCCGC 659

      78 ......ProGlyAlaSerSerArgGluGluProPro.........GluGl 89
               ||:::   |||       :::::::::|||         :::::
     660 ACAGACCCCGCCCTGTCCTGCAGCAGGAAGCCACCACTCCCCTACCACCA 709

      89 gProGluGly...............ProValIleAspSerGlnAlaP 101
         :  ||||||||             :::|||   |||   |||
     710 TGGAGAGCAGTGCCTTTACTCCAGACAAATCGCCATCAAGTCCCCCGCTC 759

     101 roAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHis 117
         ||:::           :::||:|||         |||:::
     760 CCGGT...............GCCCCTGGACAGTCGCCCCTGCAGCCCTTT 794

     118 Ser...LeuGluGlnValGlnSerMetValValGlyGluValLeuLysAs 133
         |||     ||||||   ||||||::::::   ::::::      :::::
     795 TCCAGGGCAGAACAGCAGCAGAGCCTCCTGAGAGCCTCCAGCTCTTCCCA 844

     134 pIleGluThrAlaCysLysLeuLeu............AsnIleThrAlaA 146
         :  :::   :::          |||           ::::::      :
```

```
  845 GTCCCACCCTGGCCACGGGTACCTTGGTGAGCACAGCTCCGTCTTCCAGC 894

  146 spProMetAspTrpSerProSerAsnValGluLysTrpProLeuPhe:pIle 192
      ::|||:::|||
  895 AGCCCGTGGACATGTGC.............................. 911

  163 GluHisGlnTyrArgLeuPro.................ProMetGlyLy 179
         |||:::::::      |||                |||:::
  912 ...CACTCCTTCACATCTCCTCAGGAGGGGGCCGGGAACCTCTCCCAGC 958

  173 sAlaPheGluGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluC 190
        :::|||:::      ::::::                 :::::::
  959 CCCCTATCAACACCAACTGTCGGAGCCCTGCCCACCCTACCCCCAGCAGA 1008

  190 lnPheArgGluArg..................................... 194
      ::|||:::|||:::
 1009 ACTTCAAGCAGGAGTACCATGACCCCCTGTACGAACAGGCTGGCCAGCCC 1058

  195 ...SerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLy 210
         |||      ||||||      |||              :::
 1059 GCTTCAAGCCAGGGTGGGGTCAGTGGGCACAGG.........TACCC 1096

  210 sSerAla.AlaTrpMetLysGluArg.....ThrSerProGlyAla 223
       :::||| :::|||   ::::::|||         |||||||||   :::
 1097 AGGGGCGGGGGTGGTGATCAAACAGGAGCGCACAGACTTCGCCTACGACT 1146

  224 ..............IleHisTyrCysAlaSerThrSerGluGluSer.I 235
                    :::|||   |||:::|||||:::::   ||| |
 1147 CAGATGTCCCTGGATGTGCATCAATGTACCTCCACCCAGAGGGCTTCTCT 1196

  235 rpThrAspSerGlu..................................... 239
      |||||   ||||::
 1197 GGACCCTCTCCAGGTGATGGAGTGATGGTTATGGCTATGAAAAATCCCT 1246

  240 .........Val.AspSerSerCys.......................... 244
                 ||||::   |||
 1247 TCGACCATTCCCAGATGATGTCTGCATTGTCCCTAAAAAATTTGAAGGAG 1296

  245 ..............................SerGlyGluPro...... 248
                                    :::|||   |||
 1297 ACATCAACGCAGGAAGGCATTGGAGCTTTCCGGGAGGGGCCACCCTACCAG 1346

  249 ............IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLy 261
                 ::::::|||||||||||||||||      ||||||
 1347 CGCCGGGCTGCCTTACAACTGTGGCAGTTTCTGGTGGCCCTGCTGGATGA 1396

  261 sProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyI 278
       |||   :::   ::::::||||||   |||   :::       ::::||:
 1397 CCCA...ACAAATGCTCATTTCATTGCTTGGACAGGC......CGGGGAA 1437

  278 le...PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIle 293
      ::   ||||||::   :::      :::||||||||||||||||||||
 1438 TGCAGTTTAAACTAATTGAACCTGAAGAGGTTGCCAGGCTCTGGGGTATC 1487

  294 ArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleAr 310
      :::|||||||||||||||||||||||||||||||||||||||||||:::||
 1488 CACAAGAACCGGCCAGCCATGAATTATGACAAGCTGAGCCGCTCGCTGCG 1537

  310 qGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGluArgL 327
      |   ||||||::|||||||||:::::|||          :::::::||
 1538 ATACTATTATGAGAAAGGCATCATGCAGAAGGTC...GCTGGCGAACGCT 1584

  327 euValTyrGlnPheVal 332
       ||||||:::||||||
 1585 ACGTGTACAAGTTTGTC 1601

seq_name: /equn2_6/prodata/1/1ina/6B_COMB_seq:US-09-116-049-1

seq documentation block:
; Sequence 1, Application US/09116049A
; Patent No. 6248851
; GENERAL INFORMATION:
;  APPLICANT: Hung, Mien-Chie

;  TITLE OF INVENTION: HUMAN PEA3 IS A TUMOR SUPPRESSOR FOR CANCER CELLS
;  FILE REFERENCE: UTSC:582
;  CURRENT APPLICATION NUMBER: US/09/116,049A
;  CURRENT FILING DATE: 1998-07-15
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2410
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-116-049-1


alignment_scores:
            Quality:  229.50              Length:    491
              Ratio:  1.153                 Gaps:     26
 Percent Similarity:  40.530   Percent Identity: 24.440

alignment_block:
 US-09-126-945B-2 x US-09-116-049-1   ..

 Align_seq 1/1  to: US-09-116-049-1  from: 1  to: 2410

   12 SerProSerHisLeuLeuLeuProProAspThrValSerArgThr..... 26
      ||||||:::|||         ||||||  :::   :::::::|||
  243 TCCCCCGGGCATGCTGAGCATCCCCCCCGGGCTCCGGCACAGACGCCCGC 292

   27 .........................GlyLeuGluLysAlaAlaAlaGlyA 45
                                  |||          ::::::::::|||:
  293 ACCTCAGGTCTCTGCCTCCGGCGGGGGCCCGGCCCTGTGGCCGGAGGGA 342

   45 laValGlyLeuGluArgArg.............................. 41
      ::      :::||||||||||
  343 GCGGCCGGATGGAGCGGAGGATGAAAGGCGGATACTTGGACCAGCGAGTG 392

   41 ..................................................... 41

  393 CCCTACACCTTCTGCAGCAAATCTCCCGGAAATGGAGCTTGGGCGAAGC 442

   42 ............................AspTrpSerProSerProProA 49
                                  |||   :::      ||||||:
  443 GCTGATGGTCCCGCAGGGAAAGCTCATGGACCCGGGCTCCCTGCCGCCTT 492

   49 laThrProGlu.........GlnGlyLeuSerAlaPheTyrLeuSerTyr 62
      ::      |||        |||   ||||||   |||
  493 CCGACTCAGAAGATCTCTTCCAGGACCTCAGTCACTTC............ 530

   63 PheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla..... 77
                        :::::::::::|||   ||||:|||
  531 ..................CAACAGACGTGGCTCGCAGAAGCTCAGGT 559

   77 ..................................................... 77

  560 ACCGGACAGTGATGAGCAGTTTGTTCCTGATTTCCATTCAGAAAACTTAG 609

   77 ..................................................... 77

  610 CTTTCCATAGCCCCACCACCAGGATCAAGAAGGAACCCCAGAGTCCCCGC 659

   78 ......ProGlyAlaSerSerArgGluGluProPro.........GluGl 89
           ||:::   |||      :::::||||||          :::::
  660 ACAGACCCCGCCCTGTCCTGCAGCAGGAAGCCACCACTCCCCTACCACCA 709

   89 uProGluGlnCys..............ProValIleAspSerGluAlaP 101
      :   |||||||||                 :::|||   |||   ||||
  710 TGGAGAGCAGTGCCTTTACTCCAGACAAATCGCCATCAAGTCCCCCGCTC 759

  101 roAlaGlySerLeuAspLeuValProGlyGlyLeuPheLeuGluGluHis 117
      ||:::                 :::||||||          ||:::
  760 CCGGT..............GCCCCTGGACAGTCGCCCCTGCAGCCCTTT 794

  118 Ser...LeuGluGlnValGlnSerMetValValGlyGluValLeuLysAs 133
      |||      ||||||   ||||||:::::   ::::::      :::::
```

```
  745 TCCAGGGCAGAACAGCAGCAGAGCCTCCTGAGAGCCTCCAGCTCTTCCCA 844

  133 pIleGluThrAlaCysLysLeuLeu.....AsnIleThrAlaA 146
      :   :::   :::       |||           ::::::     :
  845 TTCCCACCCTGGCCACGGGTACCTTGGTGAGCACAGCTCCGTCTTCCAGC 894

  146 spProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThr 162
      ::|||:::|||
  895 AGCCCGTGGACATGTGC..................................... 911

  163 GluHisGlnTyrArgLeuPro.....................ProMetGlyLy 178
      |||:::::::      |||                     ||||::
  912 ...CACTCCTTCACATCTGCTCAGGGAGGGGGCCGGGAACCCTCTCCCAGC 958

  179 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluG 190
      :::|||:::       ::::::                  :::::::
  959 CCCCTATCAACACCAACTGTCGGAGCCCTGCCCACCCTACCCCCAGCAGA 1008

  190 lnPheArgGlnArg..................................... 194
      ::::::::::::
 1009 ACTTCAAGCAGGAGTACCATGACCCCCTGTACGAACAGGCTGGCCAGCCC 1058

  195 ...SerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLy 210
         |||      ||||||         |||                 :::
 1059 GCTTCAAGCCAGGGTGGGGTCAGTGGGCACAGG.....TACCC 1096

  210 sSerAla AlaTrpMetLysGlnArg          ThrSerProGlyAla 223
      :::||| :::|||   :::::|||          |||||||||   :::
 1097 AGGGCCGGGGGTGGTCATCAAACAGGAAGGCACAGACTTCGGCTACGACT 1146

  224                 IleHisTyrCysAlaSerThrSerGluGluSer 235
                      ::|||   |||:::|||||:::::   ||| |
 1147 CAGATGTCCCTGGATGTGCATCAATGTACCTCCAGCCAGAGGGCTTCTCT 1196

  236 rpThrAspSerGlu..................................... 239
      |||||   ||||::
 1197 GGACCCTCTCCAGGTGATGGAGTGATGGGTTATGGCTATGAAAAATCCCT 1246

  240 .........Val.AspSerSerCys.......................... 251
                 ||||::   |||
 1247 TCGACCATTCCCAGATGATGTCTGCATTGTTCCTAAAAAATTTGAAGGAG 1296

  245 ..............................SerGlyGlnPro..... 248
                                   ::|||   |||
 1297 ACATCAAGCACGAAGGCATTGCAGCTTTCCGCCAGGGCCACCCTACCAG 1346

  249 ............IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLy 261
                 :::::|||||||||||||||||      ||||||
 1347 CGCCGGGGTGCCTTACAACTGTGGCAGTTTCTGGTGGCCCTGCTGGAGA 1396

  262 sProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyI 278
      |||   :::   ::::::::::::   |||   :::      ::::::
 1397 CGCA...ACAAATGCTCATTTCATTGCTTGGACAGGC.....CGGGGAA 1437

  278 le...PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIle 293
      ::   ||||||::   :::      :::|||||||||||||||||||||
 1438 TGGAGTTTAAACTAATTGAACCTGAAGAAGTTGCCAGGCTCTGGGGTATC 1487

  294 ArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleAr 310
      :::|||||||||||||||||||||||||||||||||||||||||||:::|
 1488 CAGAAGAACCGGCCAGCCATGAATTATGACAAGCTGAGCGGCTCGCTCGC 1537

  310 qGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgL 327
      |   ||||||::|||||||||||:::|||         ::::::|||
 1538 ATACTATTATGAGAAAGGCATCATGCAGAAGGTG...GCTGGCGAACGCT 1584

  327 euValTyrGlnPheVal 332
        ||||||::||||||
 1585 ACGTGTACAAGTTTGTG 1601

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB_seq:US-09-213-767-1

seq_documentation_block:
```

```
; Sequence 1, Application US/09213767
; Patent No. 5948680
; GENERAL INFORMATION:
;  APPLICANT: Brenda F. Baker
;  APPLICANT: Lex M. Cowsert
;  TITLE OF INVENTION: ANTISENSE MODULATION OF ELK-1 EXPRESSION
;  FILE REFERENCE: RTS-0024
;  CURRENT APPLICATION NUMBER: US/09/213,767
;  CURRENT FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
;   LENGTH: 2266
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (316)..(1602)
US-09-213-767-1


alignment_scores:
            Quality:  225.00              Length:     168
              Ratio:  2.228                 Gaps:       8
 Percent Similarity:  60.119   Percent Identity: 36.310

alignment_block:
 US-09-126-945B-2 x US-09-213-767-1   ..

 Align seq 1/1  to: US-09-213-767-1  from: 1  to: 2266

     169 ProProMetGlyLysAlaPheGluGluLeuAlaGlyLysGluLeuCysAl 185
         ||||||   |||   |||       ::::::::::::
     124 CCGCCCCCCGGGAGCTGCC      AACATTGCCAACGCCACCGGCCACGCT 167

     185 aMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValL 202
           :::   :::   |||||||::       |||::   |||   ::::
     168 ACACACAGCCCCAACTTTCAGGAGA......CCCGTCCGTGGC...CTTA 208

     202 euHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGlnArg 218
         ::::::::   |||   :::   :::|||                 :::
     209 TTTATTCCACACCTTCTGTACATCGTAGC..................GAA 240

     219 ThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerIr 235
         :::     ||||||                          :::     ||
     241 TCAACCCTGGCCCCCACTCCTCCCCATCCCTCTTTAACAGTACCCCTG 290

     235 p...........ThrAspSerGluValAspSerSerCysSerGlyGlnP 248
         |           |||          ::::|   |||
     291 GGATGGCGTCAGCACTCCCCCAGCCATGCACCCATCT.............. 327

     248 rolleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSer 264
           :::   ||||||||||||||   :::   |||||::   ::::::
     328 ..GTGACGCTGTGGCAGTTTCTGCTGCAG...CTGCTGAGAGACCAAGGC 372

     265 TyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIl 281
           |||:::::::|||   |||   :::::::::   |||   |||||::
     373 AATGGCCACATCATCTCCTGGACTTCACGGGATGGTGGTGAATTCAACGCT 422

     281 eGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgP 298
         :   ||:::   :::|||||||||||||||||||:::|||||||||:::
     423 GGTGGATGCAGAGGAGGTGGCCCGGCTGTGGGGACTACGCAAGAACAAGA 472

     298 roAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLys 314
           ||||||||||||||||||||||||:::::|||   ||||||
     473 CCAACATGAATTACGACAAGCTCAGCCGGTCCATTGCGGTACTACTATGAC 522

     315 LysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPh 331
         |||:::|||||||||||   :::   :::|||::::::|||||:::||
     523 AAGAACATCATCCGCAAGGTGAGC...GGCCAGAAGTTCGTCTACAAGTT 569

     331 eVal 332
         ||||
     570 TGTG 573
```

```
Seq_name: /cgn2_6/ptodata/1/ina/26_COMB_seq:US-09-360-779-1

seq_documentation_block:
; Sequence 1, Application US/09360779
; Patent No. 6268216
; GENERAL INFORMATION:
;  APPLICANT: Deneris, Evan S.
;  APPLICANT: Fyodorov, Dmitry V.
;  APPLICANT: Hendricks, Timothy J.
;  TITLE OF INVENTION: Reagents and Methods for the Screening of Compounds
;  TITLE OF INVENTION: Useful in the Treatment of Neurological Diseases
;  FILE REFERENCE: CASE 03828
;  CURRENT APPLICATION NUMBER: US/09/360,779
;  CURRENT FILING DATE: 1999-07-26
;  EARLIER APPLICATION NUMBER: 60/094,264
;  EARLIER FILING DATE: 1998-07-27
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;  LENGTH: 1752
;  TYPE: DNA
;  ORGANISM: Rattus norvegicus
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (112)..(1131)
US-09-360-779-1


alignment_scores:
            Quality:  202.50              Length:     291
              Ratio:  1.500                 Gaps:      17
 Percent Similarity:  46.392   Percent Identity:  40.584

alignment_block:
 US-09-126-945B-2 x US-09-360-779-1  ..

 Align_seq 1/1  to: US-09-360-779-1  from: 1  to: 1752

       44 SerProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLe 60
          ||||||  ||||||:::::|||:::         |||   :::
      153 AGCCCCCCACCCCCCTCAGCCCCATCCCCTCACAGCTCACTCCTCCAGTA 202

       60 uSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysA 77
           :::         |||                 :::|||  |||   |
      203 CACCGGCACCGGGATG.................GGCTGGGATGCA...G 231

       77 laProGlyAlaSerSerArgGluGluProProGluGluProGluGln... 92
          ||||||||                 ||||||:::::|||:::
      232 CTCCAGGA.............CCCCTCCCTCCTCACCACACCCTGGC 266

       93 CysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValPr 109
          ||||||:::               |||  |||                ||
      267 TGCCCGCTC.............CCGCCAGCCCTTGCCCGA.....CC 295

      109 oGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMet 125
          |||||||::: ||||||     |||                |||||
      296 CGGCGGCTCTACTCTTCCCTGTCACCCA.........CAGTCACCA 333

      126 ValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAs 142
          :::|||           :::  :::  :::|||     :::::
      334 CGGGCGGTTATCGGCACCCCAAGCGCAAAGCTGACGTGCCCCCCCGTGCG 383

      142 nIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpL 159
          :         :::|||          |||||||:::
      384 GTCCCCCCCATCTCCCACGGCCCAGTCCCCGGCA.............. 417

      159 euLeuTrpThrGluHisGluTyrArgLeuProProMetGlyLysAlaPhe 175

      417 ................................................. 417

      176 GlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGlnGlnPheAr 192
                                    ||||||   ::::::

      418 ..........................GCGATGAGACAGAGCGGCACCTC 440

      192 qGluArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleT 209
           |||      ||||||          :::::::::   :::|||
      441 cAG.....CCCCTG.........CTGATCAACATGTACCTACCAGATC 475

      209 rpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIleHis 225
           :::   :::   :::|||||      ||||||
      476 CCGTCGGAGATGGTCTTTTAAGCAAGGCAAGAGCCCG.......... 513

      226 TyrCysAlaSerThrSerGluGluSerTrpThrAsp...SerGlnValAs 241
                                  ||||||          |||   :::
      514 ......................AGCTGGGGGCCGCTGAGCCCTGCGGT 539

      241 pSerSerCysSerGlyGluProIleHisLeuTrpGlnPheLeuLysGluL 258
           ::::::   |||||||||   |||:::|||||||||||||||   ||||
      540 ACAGAAAGGCAGCGGGCAG...ATCCAGTTGTGGCAGTTCTACTGGAGC 586

      258 euLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 274
          |||||              :::   |||      |||   |||   :::
      587 TGCTGGCAGACCGCCGCAACGCCGGC...TGCATCGCGTGG...GAGGGC 630

      275 GluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTr 291
               |||   ||||||:::   |||      :::|||||||||   ||
      631 GGCCACGGCGAGTTCAAGCTCACCGACCCCGACGAGGTGGCGCGACGCTG 680

      291 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 308
          ||||   ||||||:::::::|||   |||||||||||||||||||||||:
      681 GGGCGAGCGCAAGAGCAACGCCCAATATGAACTACGACAAGCTAAGTCGAG 730

      308 erIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSer 324
          :::::|||   ||||||   ||:::|||:::   |||          :::
      731 CACTGCGCTACTACTACGACAAAAACATCATGAGCAAGGTGCAC...GGC 777

      325 GlnArgLeuValTyrGlnPhe 341
          :::|||   :::|||:::|||
      778 AAGCGCTACGCCTACCGCTTT 798

Seq_name: /cgn2_6/ptodata/1/iba/5B_COMB_seq:Ds-08-469-412A-1

seq_documentation_block:
; Sequence 1, Application US/08469412A
; Patent No. 5856125
;  GENERAL INFORMATION:
;     APPLICANT:  Mavrothalassitis, George J.
;     APPLICANT:  Blair, Donald G.
;     APPLICANT:  Fisher, Robert J.
;     APPLICANT:  Beal Jr., Gregory J.
;     APPLICANT:  Athanasiou, Meropi A.
;     APPLICANT:  Sgouras, Dionyssios N.
;     TITLE OF INVENTION:  The ERF Genetic Locus and Its Products
;     NUMBER OF SEQUENCES:  16
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Townsend and Townsend and Crew LLP
;       STREET:  Two Embarcadero Center, Eighth Floor
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94111-3834
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/469,412A
;       FILING DATE:  05-JUN-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Garrett Wackowski, Eugenia
;       REGISTRATION NUMBER:  37,830
;       REFERENCE/DOCKET NUMBER:  015280-229000
;     TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (415) 576-0200
;       TELEFAX:  (415) 576-0300
;    INFORMATION FOR SEQ ID NO:  1:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH:  2667 base pairs
;        TYPE:  nucleic acid
;        STRANDEDNESS:  single
;        TOPOLOGY:  linear
;      MOLECULE TYPE:  cDNA
;      FEATURE:
;        NAME/KEY:  CDS
;        LOCATION:  123..1769
;        OTHER INFORMATION:  /note= "human ERF (ETS2 Repressor
;        OTHER INFORMATION:  Factor) cDNA"
US-08-469-412A-1


alignment_scores:
            Quality:  194.50              Length:      91
              Ratio:  2.860                 Gaps:       3
 Percent Similarity:  74.725   Percent Identity: 47.253

alignment_block:
 US-09-126-945B-2 x US-08-469-412A-1    ..

 Align seq 1/1  to: US-08-469-412A-1  from: 1  to: 2667

      241 AspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysG: 257
          :::|||||:  :::::::::  ||||:::||||||:::|||:::   ||
      177 GAGTCGTCCCCCTGGCTCAAGGCAGATCCAGCTGTGGCACTTTATCCTGGA 226

      257 uLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnL 274
          |||||||  |||  :::::|||         |||  |||   :::
      227 GCTGCTGCGGAAG...GAGGAGTACCAGGGCGTCATTGCCTGG...CAGG 270

      274 ysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeu 290
            :::  |||  |||  |||:::|||      :::|||||||||||||
      271 GCGACTACCGCGGAATTCGTCATCAAAGACCCTGATGAGGTGCCCCCGCTG 320

      291 TrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerAr 307
          |||||||:|||||||  :::|||  ||||||||||||||||||||||||
      321 TGGGGCGTTCGCAAGTGCAAGCCCCAGATGAATTACGACAAGCTGAGCCG 370

      307 qSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleS 324
          |:::::::||  |||||||:|||  |||:::::|||         :
      371 GGCCCTGCGCTATTACTATAACAAGCGCATTCTGCACAAGACCAAG...G 417

      324 erGlnArgLeuValTyrGlnPhe 338
          :::::|||:::::|||:::|||
      418 GCAAACGGTTCACCTACAAGTTC 440

seq_name: /cgn2_6/ptodata/1/iin/96b_0MB/seq-US-09-021-715-1

seq_documentation_block:
; Sequence 1, Application US/09021715
; Patent No. 6194547
;   GENERAL INFORMATION:
;       APPLICANT: Mavrothalassitis, George J.
;                  Blair, Donald G.
;                  Fisher, Robert J.
;                  Beal Jr., Gregory J.
;                  Athanasiou, Meropi A.
;                  Sgouras, Dionyssios N.
;       TITLE OF INVENTION: The ERF Genetic Locus and Its Products
;       NUMBER OF SEQUENCES: 16
;       CORRESPONDENCE ADDRESS
;            ADDRESSEE: Townsend and Townsend and Crew LLP
;            STREET: Two Embarcadero Center, Eighth Floor
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94111-3834
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/021,715
;            FILING DATE: 10-Feb 1998
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Garrett-Wackowski, Eugenia
;            REGISTRATION NUMBER: 37,430
;            REFERENCE/DOCKET NUMBER: 015280-229000
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 576-0200
;            TELEFAX: (415) 576-0300
;    INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 2667 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  123..1769
;            OTHER INFORMATION:  /note= "human ERF (ETS2 Repressor
;            Factor) cDNA"
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-021-715-1


alignment_scores:
            Quality:  194.50              Length:      91
              Ratio:  2.860                 Gaps:       3
 Percent Similarity:  74.725   Percent Identity: 47.253

alignment_block:
 US-09-126-945B-2 x US-09-021-715-1    ..

 Align seq 1/1  to: US-09-021-715-1  from: 1  to: 2667

      241 AspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysG: 257
          :::|||||:  :::::::::  ||||:::||||||:::|||:::   ||
      177 GAGTCGTCCCCCTGGCTCAAGGCAGATCCAGCTGTGGCACTTTATCCTGGA 226

      257 uLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnL 274
          |||||||  |||  :::::|||         |||  |||   :::
      227 GCTGCTGCGGCAAG...GAGGAGTACCAGGGCGTCATTGCCTGG...CAGG 270

      274 ysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeu 290
            :::  |||  |||  |||:::|||      :::|||||||||||||
      271 GCGACTACCGCGGAATTCGTCATCAAAGACCCTGATGAGGTGCCCCCGCTG 320

      291 TrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerAr 307
          |||||||:|||||||  :::|||  ||||||||||||||||||||||||
      321 TGGGGCGTTCGCAAGTGCAAGCCCCAGATGAATTACGACAAGCTGAGCCG 370

      307 qSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleS 324
          |:::::::||  |||||||:|||  |||:::::|||         :
      371 GGCCCTGCGCTATTACTATAACAAGCGCATTCTGCACAAGACCAAG...G 417

      324 erGlnArgLeuValTyrGlnPhe 338
          :::::|||:::::|||:::|||
      418 GCAAACGGTTCACCTACAAGTTC 440

seq_name: /cgn2_6/ptodata/1/iin/96b_0MB/seq-US-08-306-691B-50

seq_documentation_block:
; Sequence 50, Application US/08306691B
; Patent No. 5734039
;   GENERAL INFORMATION:
;     APPLICANT:  Calabretta, Bruno
;     APPLICANT:  Skorski, Tomasz
```

```
;     TITLE OF INVENTION:  ANTISENSE
;     TITLE OF INVENTION:  OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
;     NUMBER OF SEQUENCES:  55
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Seidel, Gonda, Lavorgna & Monaco, P.C.
;       STREET:  Two Penn Center, Suite 1800
;       CITY:  Philadelphia
;       STATE:  Pennsylvania
;       COUNTRY:  U.S.A.
;       ZIP:  19102
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 720 Kb
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  MS DOS
;       SOFTWARE:  WordPerfect 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/306,691B
;       FILING DATE:  September 15, 1994
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Monaco, Daniel A.
;       REGISTRATION NUMBER:  30,480
;       REFERENCE/DOCKET NUMBER:  8321-8
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (215) 568-8383
;       TELEFAX:  (215) 568-5549
;       TELEX:  No. 5734039
;   INFORMATION FOR SEQ ID NO:  50:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1364 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-306-691B-50


alignment scores:
            Quality:  170.00              Length:     359
              Ratio:  0.988                 Gaps:      21
 Percent Similarity:  47.911   Percent Identity: 25.905

alignment block:
 US-09-126-945B-2 x US-08-306-691B-50   ..

 Align seq 1/1  to: US-08-306-691B-50  from: 1  to: 1364

       20 ProAspThrValSerArgThrGlyLeuGluLysAla...AlaAlaGlyAl 35
          |||   :::|||:::      |||      :::|||   :::|||   ::
       22 CCCTGCAATGTCAAGGGAGGGGGCTCACCCAGGGCTCCTGTAGCTCAGGG 71

       36 aValGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProG 52
          :::|||||||                |||||:::|||        |||:
       72 GGCAGGCCTG..............AGCCCTGCACCCGCCCCACGACCGT 106

       52 IuGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrPro 68
          :::::        |||                              :::|||
      107 CCAGCCCCTGACGGGCA......................CCCCATCCT 132

       69 Glu..AspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGlnG 85
          |||    |||:::   |||            |||      :::   :::|
      133 GAGGGACTCTGCATTGGCCCCCACCGAGGCAGGGGATCTGACCGACTCGG 182

       86 IuProProGlnGluProGluGlnCys..............ProValIle 96
          |||||           :::   |||                |||||:::
      183 AGCCCGGCTGGATGTTACAGGCCTGCAAAATGGAAGGGTTCCCCTCGTC 232

       97 AspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuTh 113
                     |||      |||   ||||||||||||
      233 ..........CCCCCTCCATCAGAAGACCTGGTGCCCTATGACACCGA 270

      113 rLeu.........GluGluHis.SerLeuGluGlnValGlnSerMetVal 126
           |||          :::   |||.|||:::      :::   |||   |||
      271 TCTATACCAACGCCAAACGCACGAGTATTACCCCTATC...TCAGCAGTG 317

      127 ValGlyGluValLeuLysAspIle.......................... 134
          :::|||:::::::::::      |||
      318 ATGGGCAGAGCCATAGCGACCATTACTGGGACTTCCACCCCCACCACGTG 367

      135 ..........................................GluThrA 147
                                                     :::   |
      368 CACAGCGAGTCCGAGAGCTTCGCCGAGAACAACTTCACGGAGCTCCAGAG 417

      147 IaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSer 163
          |||||:::      :::   :::::::::          ||||||
      418 CGTGCAGCCCCCGCAGCTGCAGCAGCTCTACCGCCACATGGACC..... 461

      164 AsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProPr 180
                               |||:::::::              :::|||||
      462 .....................TGGAGCAGATGCACGTCCTCGATACCCCC 490

      170 oMetGlyLysAlaPheGluGluLeuAlaGlyLysGluLeu.......... 183
          |           :::       ||||||   :::::
      491 ATGGTGCCACCCCATCCCAGTCTTGGCCACCAGGTCTCCTACCTGCCCCG 540

      184 ..CysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGly 199
            ||||||   |||   :::          |||   ||||||
      541 GATGTGCCTCCAGTACCCATCCCTGTCCCCAGCCCAGCCCA......... 581

      200 AspValLeuHisAlaHisLeuAspIleTrpLysSerAlaIaTrpMetLy 216
                                                |||   ||||:
      582 ..........................................GCTCAGATGAG 592

      216 sGluArgThrSerProGlyAlaIleHisTyr...CysAlaSerThrSerG 232
          :::|||:::|||      |||   ||||::   |||   ::::::   :
      593 GAGGAGGGCGAGCGGCAGAGCCCCCCACTGGAGGTGTCTGACGGCGAGGC 642

      232 IuGluSerTrpThrAspSerGluValAspSerSerCysSerGly...... 246
          ::   :::||||::   :::          :::||||||   |||
      643 GGATGGCCTGGAGCCCGGGCCTG......GGCTCCTGCCTGGGGAGACAG 686

      247 ......Gln..ProIleHisLeuTrpGluPheLeuLysGluLeuLeuLeu 260
                :::     |||:::|||:::||||||||||   :::   ||||||
      687 GCAGCAAGAAGAAGATCCGGCTGTACCAGTTCCTGTTGGAC...CTGCTC 733

      261 LysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGl 277
          :::      :::             |||   |||:::::|||:::||||||
      734 CGCAGCGGCGACATCAAGGACAGCATCTGGTGGGTGGACAAGGACAACGG 783

      277 yIlePheLysIleGluAspSerAlaGlu......ValAlaArgLeuTrpG 292
          |   |||||||||||||||||   :::       :::|||:::   ||||
      784 CACCTTCTACTTCTCGTCCAAGCACAAGGAGGCGCTGCCGCACCGCTGGG 833

      292 lyIleArgLys...AsnArgProAlaMetAsnTyrAspLysLeuSerArg 307
          |||||:::|||   ||||||       |||:::|||:::|||:::::|||
      834 GCATCCAGAAGGGCAACCGCAAGAAGATGACCTACCAGAAGATGGCGCGC 883

      308 SerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSe 324
          :::::::::::|||   |||   |||   :::::::       :::::
      884 GCGCTGCGCAACTACGGCAAGACGGGCGAGGTCAAGAAG......CTGAA 927

      324 rGlnArgLeuValTyrGluPhe 341
          :::::::|||:::|||||||||
      928 GAAGAAGCTCACCTACCAGTTC 949

seq_name: /cm2/6/pub/data/US/09/09126/seq/945b-2.rni 0.33 99251 65

seq_documentation_block:
; Sequence 65, Application PC/US9306251
;  GENERAL INFORMATION:
;     APPLICANT:  Wickstrom, Eric and Rife, Jason P.
;     TITLE OF INVENTION:  Trivalent Synthesis of Oligonucleotides Containing
;     TITLE OF INVENTION:  Stereospecific Alkylphosphonates and Arylphosphonates
;     NUMBER OF SEQUENCES:  98
```

```
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;       STREET:  400 Garden City Plaza
;       CITY:  Garden City
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  11530
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/06251
;       FILING DATE:  19930630
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  DiGiglio, Frank S.
;       REGISTRATION NUMBER:  31,346
;       REFERENCE/DOCKET NUMBER:  8586
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  516-742-4343
;       TELEFAX:  516-742-4366
;       TELEX:  230 901 SANS UR
;   INFORMATION FOR SEQ ID NO:  65:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1364 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
PCT-US94-06251 65


alignment_scores:
            Quality:  170.00              Length:     359
              Ratio:  0.988                 Gaps:      21
 Percent Similarity:  47.911   Percent Identity: 25.905

alignment_block:
 US-09-126-945B-2 x PCT-US94-06251-65   ..

 Align seq 1/1  to: PCT-US94-06251-65  from: 1  to: 1364

      20 ProAspThrValSerArgThrGlyLeuGluLysAla...AlaAlaGlyAl 45
         |||  :::|||:::      |||      :::|||  :::|||  ::
      22 CCCTGCAATGTCAAGGGAGGGGGCTCACCCAGGGCTCCTGTAGCTCAGGG 71

      46 aValGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProG 92
         :::||||||              |||||||:::|||         |||:
      72 GGCAGGCCTG..............AGCCCTGCACCCGCCCCACGACCGT 106

      93 lnGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrPro 98
         :::::      |||                              :::|||
     107 CCAGCCCCTGACGGGCA.......................CCCCATCCT 132

      99 Glu.AspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluG 85
         |||   |||:::  |||                |||     :::   ::::
     133 GAGGGGCTCTGCATTGGCCCCCACCGAGGCAGGGGATCTGACCGACTCGG 182

      85 lnProProGluGluProGluGlnCys..............ProValIle 96
         |||||            :::   |||               |||||||
     183 AGCCCGGCTGGATGTTACAGGCGTGCAAAATGGAAGGGTTTCCCCTCGTC 232

      97 AspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuTh 113
                 |||       |||   ||||||||||||
     233 ............CCCCCTCCATCAGAAGACCTGGTGCCCTATGACACGGA 270

     113 rLeu.........GluGluHis.SerLeuGluGlnValGlnSerMetVal 126
          |||          :::   ||| |||:::      :::   |||   |||
     271 TCTATACCAACCGCCAAACGCACGACTATTACCCCTATC....TCAGCACTG 317

     127 ValGlyGluValLeuLysAspIle.......................... 134

         :::|||:::::::::      |||
     318 ATGGGCAGACCCATAGGGACCATTACTGGGACTTCCACCCCCACCACGTG 367

     135 ..........................................GluThrA 147
                                                    :::  |
     368 CACAGCCACTTCGAGAGCTTCGCCGAGAACAACTTCACGGAGCTCCAGAG 417

     137 laCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSer 153
         ||||:::      :::   :::::::::          ||||||
     418 CGTGCAGCCCCCGCAGCTGCAGCAGCTCTACCGCCACATGGAGC...... 461

     154 AsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProPr 170
                            |||::::::          :::||||||
     462 ..................TGGAGCAGATGCACGTCCTCGATACCCCC 490

     170 oMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeu.......... 188
         |          :::      ||||||  ::::::
     491 ATGGTGCCACCCCATCCCAGTCTTGGCCACCAGGTCTCCTACCTGCCCCG 540

     184 ..CysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGly 199
           ||||||   |||   :::          |||   ||||||
     541 GATGTGCCTCCAGTACCCATCCCTGTCCCCAGCCCAGCCCA......... 581

     200 AspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLy 216
                                                 |||  ||:::
     582 ..................................GCTCAGATGAG 592

     216 sGluArgThrSerProGlyAlaIleHisTyr...CysAlaSerThrSerG 242
         :::|||:::||       |||   |||:::   |||   ::::::   :
     593 CAGGAGGGCGAGCGGCAGAGCCCCCCACTGGAGGTGTCTGACGGCGAGGC 642

     232 luGluSerTrpThrAspSerGluValAspSerSerCysSerGly...... 246
         ::   :::|||:::   :::          :::|||||   |||
     643 GGATGGCCTGGAGCCCGGGCCTG......GGCTCCTGCCTGGGGAGACAG 686

     247 ......Gln..ProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeu 260
               :::    |||:::|||:::|||||||||   :::   ||||||
     687 GCAGCAAGAAGAAGATCCGCCTGTACCAGTTCCTGTTGCAC...CTGCTC 733

     261 LysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGl 277
         :::    :::          |||  |||:::::|||:::||||||
     734 CGCAGCGGCGACATGAAGGACAGCATCTGGTGGGTGGACAAGGACAAGGG 783

     277 yIlePheLysIleGluAspSerAlaGln......ValAlaArgLeuTrpG 292
         |  |||:::::::::::::::::   :::      :::|||:::   ||||
     784 CACCTTCCAGTTCTCGTCCAAGCACAAGGAGGCGCTGGCGCACCGCTGGG 833

     292 lyIleArgLys...AsnArgProAlaMetAsnTyrAspLysLeuSerArg 307
         ||||:::::||   ||||||      |||:::|||:::|||:::||||
     834 GCATCCAGAAGGGCAACCGCAAGAAGATGACCTACCAGAAGATGGCGCGC 883

     308 SerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSe 324
         :::::|||:::|||   |||   |||   :::::|||      :::::
     884 GCGCTGCGCAACTACGGCAAGACGGGCAGGTCAAGAAG......GTGAA 927

     324 rGlnArgLeuValTyrGlnPhe 344
         :::::::|||:::||||||||
     928 GAAGAAGCTCACCTACCAGTTC 949

Seq_name: "no:1_pf:dir:1:0:5B,COMB.seq:US-08-469-412A-6

seq_documentation_block:
; Sequence 6, Application US/08469412A
; Patent No. 5856125
;  GENERAL INFORMATION:
;     APPLICANT:  Mavrothalassitis, George J.
;     APPLICANT:  Blair, Donald G.
;     APPLICANT:  Fisher, Robert J.
;     APPLICANT:  Beal Jr., Gregory J.
;     APPLICANT:  Athanasiou, Meropi A.
;     APPLICANT:  Sgouras, Dionyssios N.
;     TITLE OF INVENTION:  The ERF Genetic Locus and Its Products
;     NUMBER OF SEQUENCES:  16
```

```
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Townsend and Townsend and Crew LLP
;       STREET:  Two Embarcadero Center, Eighth Floor
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94111-3834
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/469,412A
;       FILING DATE:  05-JUN-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Garrett-Wackowski, Eugenia
;       REGISTRATION NUMBER:  37,330
;       REFERENCE/DOCKET NUMBER:  015280-229000
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 576-0200
;       TELEFAX:  (415) 576-0300
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2544 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:
;       LOCATION:  1..2544
;       OTHER INFORMATION:  /note= "genomic DNA sequence for murine
;       OTHER INFORMATION:  ERF gene (promoter region and first exon
;       OTHER INFORMATION:  not included)"
US-08-469-412A-6


alignment scores:
            Quality:  160.00              Length:      70
              Ratio:  4.077                 Gaps:       2
 Percent Similarity:  74.286   Percent Identity: 50.000

alignment block:
 US-09-126-945B-2 x US-08-469-412A-6   ..

 Align seq 1/1  to: US-08-469-412A-6  from: 1  to: 2544

      241 AspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGl 287
          :::|||||||   ||||||||||   |||:::|||||||:::|||:::   ||
      417 GAGTCATCCCCTGGCTCCAGGCAGATCCAGCTGTGGCACTTTATCCTGGA 466

      287 nLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnL 274
          |||||||   |||   :::|||:|||         |||   |||   :::
      467 GCTGCTTCGGAAA...GAGGAGTACCAGGGCGTCATCGCTTGG...CAGG 510

      274 ysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeu 290
          :::   |||   |||   |||:::|||       :::|||||||||||||
      511 GGCACTACGGGGAGTTTGTCATCAAGGACCCTGATGAAGTGGCTCGCCTC 560

      291 TrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerAr 307
          |||||:::|||||   :::|||   |||||||||||||||||||||||||
      561 TGGGGGTCCGCAAGTGCAAACCCCAGATGAACTATGACAAGCTCAGCCG 610

      307 qSerIleArg 410
          |:::::|||
      611 GGCTTTGCGG 620

Seq name: /cpv_6/pl=data_disc5:=cMb.seq;US-09-021-715-6

Seq documentation block:
; Sequence 6, Application US/09021715
```

```
; Patent No. 6194547
;   GENERAL INFORMATION:
;       APPLICANT: Mavrothalassitis, George J.
;                  Blair, Donald G.
;                  Fisher, Robert J.
;                  Beal Jr., Gregory J.
;                  Athanasiou, Meropi A.
;                  Sgouras, Dionyssios N.
;       TITLE OF INVENTION: The ERF Genetic Locus and Its Products
;       NUMBER OF SEQUENCES: 16
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Townsend and Townsend and Crew LLP
;            STREET: Two Embarcadero Center, Eighth Floor
;            CITY: San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94111-3834
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC DOS/MS DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/021,715
;            FILING DATE: 10-Feb-1998
;            CLASSIFICATION: Unknown
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Garrett-Wackowski, Eugenia
;            REGISTRATION NUMBER: 37,330
;            REFERENCE/DOCKET NUMBER: 015280-229000
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 576-0200
;            TELEFAX: (415) 576-0300
;   INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 2544 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
;            NAME/KEY:
;            LOCATION: 1..2544
;            OTHER INFORMATION: /note= "genomic DNA sequence for murine
;            ERF gene (promoter region and first exon
;            not included)"
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-021-715-6


alignment_scores:
            Quality:  160.00              Length:      70
              Ratio:  4.077                 Gaps:       2
 Percent Similarity:  74.286   Percent Identity: 50.000

alignment block:
 US-09-126-945B-2 x US-09-021-715-6   ..

 Align seq 1/1  to: US-09-021-715-6  from: 1  to: 2544

      241 AspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGl 287
          :::|||||||   ||||||||||   |||:::|||||||:::|||:::   ||
      417 GAGTCATCCCCTGGCTCCAGGCAGATCCAGCTGTGGCACTTTATCCTGGA 466

      287 nLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnL 274
          |||||||   |||   :::|||:|||         |||   |||   :::
      467 GCTGCTTCGGAAA...GAGGAGTACCAGGGCGTCATCGCTTGG...CAGG 510

      274 ysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeu 290
          :::   |||   |||   |||:::|||       :::|||||||||||||
      511 GGCACTACGGGGAGTTTGTCATCAAGGACCCTGATGAAGTGGCTCGCCTC 560

      291 TrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerAr 307
```

```
      ||||||:::|||||| :::||| ||||||||||||||||||||
      561 TGGGGGGTCCGCAAGTGCAAACCCCAGATGAACTATGACAAGCTGAGCCG 610

      307 qSerIleArg 310
          |:::::|||
      611 GGCTTTGCGG 620

seq_name: /tmp2_5/pub/dirty/US-08-706-292-1.rni MB seq US-08-706-292-1

seq_documentation_block:
; Sequence 1, Application US/08706292
; Patent No. 5705380
;  GENERAL INFORMATION:
;     APPLICANT:  NORTH, Micheal
;     TITLE OF INVENTION:  GENES ASSOCIATED WITH RETINAL
;     TITLE OF INVENTION:  DYSTROPHIES
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;       STREET:  4 Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94111-4187
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  08/08/706,292
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SHERWOOD, Pamela J.
;       REGISTRATION NUMBER:  36,677
;       REFERENCE/DOCKET NUMBER:  A-27889/ECS
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-494-8700
;       TELEFAX:  415-494-8771
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1733 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-706-292-1


alignment_scores:
            Quality:  108.50              Length:     814
              Ratio:  0.764                Gaps:      17
 Percent Similarity:  45.223   Percent Identity:  22.611

alignment_block:
 US-09-126-945B-2 x US-08-706-292-1  ..

 Align seq 1/1  to: US-08-706-292-1  from: 1  to: 1733

       31 AlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPr 47
          :::|||:::|||||              :::|||  |||:::|||  ||
      273 TCTGCGGGAGGAGCGCCTTTTAGGTGACAGAGGCCTGGGAACCCTTCC 322

       47 oProAlaThrProGluGluGlyLeuSerAlaPheTyrLeuSerTyrPheA 64
          ||||             :::|||  |||||:::
      323 TCGGCAAGAAAGTGTCAGAGGCACATCTGCCCTC.............. 356

       64 spMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGly... 79
                           |||      ::::::|||||
      357 .....TGGCATCCACAGTGCCCTGGGCAC 380

       80 .....AlaSerSerArgGluGluProProGluGluPr 90
                          :::::::||  :::|||||:::
      381 CGTGAGCTGTGGTGGAGACGGCAGGGCGAGCGCGGCCTCCCGACACCGC 430

       90 oGluGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspL 107
          ::::::     |||  :::|||:::::::::::     |||   |
      431 GGACAGAAGCAGTGTTCAGGAATCTGGGTCTCCAGTCCCCTTTCTTATCC 480

      107 eu.ValProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGl 123
          || :::|||  :::         |||         ||||||::|||
      481 TGGCTCCCAGACAATTCCGATGCAGAA.........TTGGAGCAAGTC.. 519

      123 nSerMetValValGlyGluValLeuLysAspIleGluThrAlaCysLysL 140
                              ::::::|||:::::
      520 ........................TCCGTGGAGAATGGT..... 544

      140 euLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGln 156
          :::::::::    ||:::   ||||||        :::
      545 .....TCCGTCTCTCCCCCACCTTTAAAACAGTCTCCGAGAATCCGACGC 579

      157 LysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLy 173
          |||        |||     ||||||          |||     ::
      580 AAGGGT.....TGGCAAGCCCACCAA.........CGACCTGGGACCCG 614

      173 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluG 190
          :|||     |||     ::::::::::::            ::::::
      615 TGCAGAGGTCAGAGTGACTCTCAGGATATGGGAGATGCACACAAGTCAC 664

      190 lnPheArgGlnArgSerPro...LeuGlyGlyAspValLeuHisAlaHis 205
                    ::|||   :::   ||||||   ::::::   :::
      665 CTAA.ACGCAC.AAAGGG.GAAL.GA.GGGA.CTCTCTATAT.AAAAC 714

      206 LeuAspIleTrpLysSerAlaAlaTrpMetLysGluArg..... 218
          |||  :::   ||:::          ||||:::|||
      715 TTGGCCTTCCAAAACGAAGAACACTTGCAAAACAAGAGAGGCCTCTGA 764

      219 .....hr.rProC,AlaIleHi.....fpCysAlaS 229
                ||:::   :::|||   |||           :
      765 GTCTACAGGGACGAACTCCTCAGCAGCACACAACGAAGAGTTGTCCAAGG 814

      229 crThrSerGluGlnSerTrpThrAspSerGln..... 239
          ::   :::   |||:::   |||||||||:::
      815 CCCTGAAAGGCGAGGGTGGCACGGACAGCGACCATATGAGGCACGAAGCC 864

      240 .........ValAspSerSerCysSerGlyGlnProIleHisLeuTrpGl 253
                  :::   |||   |||   |||           |||   ::
      865 TCCTTGGCAATCCGCTCCCCCTGCCCTGGG.....CTGGAGGA 902

      253 nPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleA 270
          :  ::::::    :::|||:::|||          |||   :::::::
      903 GGACATGGAAGCCTACGTGCTGCGGCCAGCGCTCCCGGGCACCATGATGC 952

      270 rq...TrpLeuAsnLys.............GluLysGlyIlePheLys 280
          ::   ::::|||:::::                 :::|||||:::|||
      953 ACTGCTACCTCTACCCGTGACAAGCACGGCGTGGACAAGGGCTTGTTCCCC 1002

      281 Ile.....................GluAspSerAlaGlnValAlaArgLe 290
                                  :::|||||   |||         ||
     1003 CTCTACTACCTCTACCTGGAGACCTCTGACAGCCTGCAGCGCTTCCTCCT 1052

      290 uTrpGlyIleArgLysAsnArg...ProAlaMetAsnTyr 302
          |  |||   :::::::::|||      :::   ||||||
     1053 GGCTGGGCGAAAGAGAAGAAGGAGCAAAACTTCTAATTAC 1092

seq_name: /tmp2_5/pub/dirty/US-09-042-365A.rni MB seq US-09-042-365A-14

seq_documentation_block:
; Sequence 14, Application US/09042365A
; Patent No. 6114502
;  GENERAL INFORMATION:
;     APPLICANT:  No, 6114502th, Michael
;     APPLICANT:  Nishina, Patsy
;     APPLICANT:  Naggart, Juergen
```

```
;     APPLICANT:  No. 6114502en Trauth, Konrad
;     TITLE OF INVENTION:  GENE FAMILY ASSOCIATED WITH
;     TITLE OF INVENTION:  NEUROSENSORY DEFECTS
;     NUMBER OF SEQUENCES:  67
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Bozicevic & Reed, LLP
;       STREET:  285 Hamilton Avenue, Suite 200
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94301
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/032,365A
;       FILING DATE:
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sherwood, Pamela J
;       REGISTRATION NUMBER:  36,677
;       REFERENCE/DOCKET NUMBER:  SEQ 2CIP2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650 327 3400
;       TELEFAX:  650 327 3231
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1733 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-09-032-365A-14


Alignment scores:
              Quality:  108.50              Length:     314
                Ratio:  0.764                Gaps:      17
 Percent Similarity:  45.223   Percent Identity: 22.611

Alignment block:
 US-09-126-945B-2 x US-09-032-465A-14   ..

 Align seq 1/1  to: US-09-032-465A-14  from: 1  to: 1733

       41 AlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPr 47
          :::|||:::|||||||             :::|||   |||:::|||   ||
      274 TCTGCCGGGAGGAGCGCCTTTTAGGTGACAGAGGCCTTGGGAACCCTTTCC 322

       47 oProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheA 64
          ||||            :::|||   |||||||:::
      323 TCCGGAAGAAAGTGTCAGAGGCACATCTGCCCTC.............. 356

       64 spMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGly... 79
                                  |||      ::::::|||||||
      357 .........................TGGCATCCACAGTGCCCTGGGCAC 380

       80 ..................AlaSerSerArgGluGluProProGluGluPr 90
                           :::::::|||   :::|||||||:::
      381 CGTGAGCTGTGGTGGACACGGCAGGGGCGAGCGCGGCCTCCCGACACCGC 430

       90 oGluGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspL 107
          ::::::      |||   :::|||:::::|||:::      |||   |
      431 GGACAGAAGCAGTGTTCAGGAATCTCGGTCTCCAGTCCCTTTCTTATCC 480

      107 euLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGluGluValGl 123
          || :::|||   :::         |||         |||||||:::|||

      481 TGGCTCCCAGACAATTCCGATGCAGAA.........TTGGAGGAAGTC.. 519

      123 nSerMetValValGlyGluValLeuLysAspIleGluThrAlaCysLysL 140
                                        :::::::|||:::::::
      520 ..............................TCCGTGGAGAATGGT....... 544

      140 euLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGln 156
          :::::::::       |||:::       ||||||          :::
      545 .....TGGGTCTCTCCCCCACCTTTTAAACAGTCTCCCAGAATCCGACGC 579

      157 LysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLy 173
          |||          |||      ||||||            |||      ::
      580 AAGGGT......TGGCAAGCCCACCAA.........CGACCTGGGACCCG 614

      173 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluG 190
          :|||      |||       ::::::::::::                ::::::
      615 TGCAGAGGGTGAGAGTGACTCCCAGGATATGGGAGATGCACACAACTCAC 664

      190 lnPheArgGlnArgSerPro...LeuGlyGlyAspValLeuHisAlaHis 205
                    :::|||   :::   ||||||   ::::::   :::
      665 CCAATATGGCACCAAACCCTGGAATGGATGGTGACTGTGTATATGAAAAC 714

      206 LeuAspIleTrpLysSerAlaAlaTrpMetLysGluArg.......... 218
          |||   :::   |||:::             |||:::|||
      715 TTGGCCTCCAAAAGGAAGAAGACTTGGAAAAGAAGAGAGAGGCCTCTGA 764

      219 .........ThrSerProGlyAlaIleHis.........TyrCysAlaS 229
                  |||:::   :::|||   |||          :
      765 GTCTACAGGGACGAACTCCTCAGCAGCACACAACGAAGAGTTGTCCAAGG 814

      229 erThrSerGluGluSerTrpThrAspSerGlu................... 249
          ::   :::   |||:::   ||||||||||:::
      815 CCCTGAAAGCCGAGGGTGGCACGGACAGCGACCATATGAGGCACGAAGCC 864

      240 .........ValAspSerSerCysSerGlyGluProIleHisLeuTrpGl 253
                  :::   |||   |||   |||             |||   ::
      865 TCCTTGGCAATCCGCTCCCCCTGCCCTGGG.............CTGGAGGA 902

      253 nPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleA 270
          :   :::::::      :::|||:::|||          |||   :::::::
      903 GGACATGGAAGCCTACGTGCTGCGGCCAGCGCTCCCGGGCACCATGATGC 952

      270 rg...TrpLeuAsnLys..............GluLysGlyIlePheLys 280
          ::   :::|||::::::              :::||||||:::|||
      953 AGTGCTACCTCACCCGTGACAAGCACGGCGTGGACAAGGGCTTGTTCCCC 1002

      281 Ile.....................GluAspSerAlaGlnValAlaArgLe 290
          :::                  :::||||||   |||          ||
     1003 CTCTACTACCTCTACCTGGAGACCTCTGACAGCCTGCAGCGCTTCCTCCT 1052

      290 uTrpGlyIleArgLysAsnArg...ProAlaMetAsnTyr 302
          |   |||   ::::::::::|||      :::   ||||||
     1053 GGCTGGGCGAAAGAGAAGAAGGAGCAAAACTTCTAATTAC 1092

Seq_name: Sequence 9, pt-Data_Files_5b_CoMB.seq:US-08-466-589-9

Seq_documentation_block:
; Sequence 9, Application US/08466589
; Patent No. 5837489
;  GENERAL INFORMATION:
;     APPLICANT:  Elliot, Kathryn J.
;     APPLICANT:  Ellis, Steven B.
;     APPLICANT:  Harpold, Michael M.
;     TITLE OF INVENTION:  HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
;     TITLE OF INVENTION:  RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
;     NUMBER OF SEQUENCES:  12
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Brown, Martin, Haller & McClain
;       STREET:  1660 Union Street
;       CITY:  San Diego
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  92101-2926
```

```
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/466,589
;       FILING DATE:  June 5, 1995
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/028,931
;       FILING DATE:  March 8, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seidman, Stephanie L.
;       REGISTRATION NUMBER:  33,779
;       REFERENCE/DOCKET NUMBER:  6362-9950
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-238-0999
;       TELEFAX:  619-238-0062
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2450 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  both
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  267..1775
US-08-466-589-9


alignment_scores:
            Quality:  104.50              Length:     374
              Ratio:  0.661                Gaps:      20
 Percent Similarity:  42.246   Percent Identity:  22.460

alignment_block:
 US-09-126-945B-2 x US-08-466-589-9/rev   ..

 Align seq 1/1  to reverse of: US-08-466-589-9  from: 1  to: 2450

       6 ProGlyLeuSerSerValSerProSerHisLeu..................  16
         |||||||:||||||         ||||||:::|||
    2244 CCAGGCCTCTCTCTCCCCTCCCTTCAGAGCTACCATTGTGTCAGGCCAG 2195

      17 ................LeuLeuProProAspThrValSerArgThrGlyL  28
                        |||||||||          :::|||   |||   |
    2194 CCCGGTGCGGTGTGAGCTCCTGCCCTTCCACCCCTTGTCCTATACCCTC 2145

      28 euGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSer  44
         ||              :::||:::                 |||:::
    2144 TTCTCCCATGACTTGGAGGCAGC..............AGACAATCCTGT 2110

      45 ProSerPro.........ProAlaThrProGluGlnGly...........  54
         |||   |||          |||:::|||||||:::   |||
    2109 CCCCTTCCTGGCCCCTTGCCCGTCACTCCTCAGGGTGGATGGCTCTGCCA 2060

      55 ............LeuSerAlaPhe.......................I  59
                    ||||||:::::                        :
    2059 TCTCTGTCTTCCACTTAGCAGCCTTGCTCCCCTCCTCAACACTATTGCCT 2010

      59 yrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAla  75
         ::   :::       |||         |||:::::|||||||||:::::
    2009 GGGCTGGTACAAGACATGGACAGCAGCCAAAACAAAGTAGTTGGTCCAGC 1960

      76 LysAlaProGlyAlaSerSer.........ArgGluGluProProGluGl  89
                 ||||||::          ||||::
    1959 CTCCACGTTGGAGCTGCTGTGTGTGGATGGAGGCGTGGGGGAGGAGGGAC 1910

      89 uProGluGlnCysPro......ValIleAspSerGlnAlaProAlaGlyS 104
         |||:::::::::         :::      ::::::::|||||

    1909 TCCAGATGAAGCAAAAGGATGGATCTGTGGGGGCAGCGCCCCT..... 1867

     104 erLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGlu 120
                  :::|||||||:::      ::::::
    1866 ..........CTTCCTGGTAGCTCACTCGTCCATCCTCCACCCAACACT 1820

     121 GlnValGlnSerMetValValGly            GluValLeuLysAs 133
         |||||||::   |||:::|||              |||:::   |||
    1827 ACTGTGCAGCAGAGGGTGGCAGGCGTGAAAGAGCATGGAGATGAGGAAGGG 1778

     133 pIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspI 150
         :::      :::    |||   ::::::::::::              |
    1777 CCTCACTTGGAGCTGAGGGCTGAGTGGTCTGAGTGGAGCAAGGTGGTGGT 1728

     150 rpSerProSerAsnValGln.............LysTrpLeuLeuTrp 161
         ||      |||:::   :::             :::|||
    1727 GGTGTAGTTCTGGAAGAGAGGCTGCAGGAACATGCCGATGGTGC..... 1684

     162 ThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGluLe 178
         :::|||:::   |||                 |||:::::   ::::::
    1683 CAAAGACACACACAAAGACAAAGATCCAGAGGAAGAGGCGGTCGATCACC 1634

     178 uAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgS 195
             :::                 |||:::::::      |||   ||||
    1633 ATGGCGACGTACTTCCAGTCCTCACTCACGCTCTGGTCATCGTCCTCGCT 1584

     195 erProLeuGly.GlyAspValLeuHisAlaHisLeuAspIleTrpLysSe 211
         ||      |||   |||      ||||||
    1583 CCGCATGTGCPCTGCGATGAAGCGCACGCC..................... 1554

     211 rAlaAlaTrpMetLysGluArgThr.................SerProG 222
          :::      :::   |||   |||                 :::|||||
    1553 .GTCCACCGCCTCCCGGAGGCCACAGCCACACGGCTCCCCTGAGCGCCCG 1505

     222 lyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSer 248
         |||||   |||:::|||       :::   |||
    1504 GGGCCCGCCACTGCTGCACGCTCAGCCCCCAAGGC ............... 1470

     239 GluValAspSerSerCysSerGlyGlnProIleHisLeu............ 251
                              ||||||||||:::|||
    1469 ................CCCGGCCAACCCCTGCACCGACGCGCGGTTGA 1438

     252 ........................TrpGlnPheLeuLysGlu........ 257
                                 |||   |||   :::|||
    1437 CGAAGCACGTGCAGGAGTCGGCCCCTGGGGCTTCGCGCAAGAAGAGGGCT 1388

     258 ................LeuLeuLeuLysProHisSerTyr.......... 265
                         |||   :::|||::::::
    1387 CCAGCGCCCTCGCGCTCACGCTGGCGTCGCCGCAGGCGCAGGCGCTGACG 1338

     266 GlyArgPheIleArgTrpLeu......AsnLysGluLysGlyIlePheLy 280
         |||   ::::::   ||||||      :::::::::::|||   :::
    1337 GGCGCAATGATGGCGTGGCTGCTGCATGAAGAGCAGCGCGGGCAGCTTCT 1288

     280 sIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnA 297
         :|||||:::   :::   :::|||      ||:::|||   :::|
    1287 CCAGGAAGACGACCTTCACCCAGGGCGC...CATGGTGTGCGTGGTGGGC 1241

     297 rgProAlaMetAsnTyrAsp 308
         ||   :::::::::   :::
    1240 GAGCGGTGGTGCACGTTGAG 1221

seq_name= /tmp/_fpl_tmp/3/fcs/SP_COMP seq:US 08-700-636-9

seq_documentation_block:
; Sequence 9, Application US/08700636
; Patent No. 5910582
;  GENERAL INFORMATION:
;     APPLICANT:  Elliot, Kathryn J.
;     APPLICANT:  Ellis, Steven B.
;     APPLICANT:  Harpold, Michael M.
;     TITLE OF INVENTION:  HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
;     TITLE OF INVENTION:  RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
```

```
;     NUMBER OF SEQUENCES:  12
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;       STREET:  444 South Flower Street, Suite 2000
;       CITY:  Los Angeles
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  90071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/700,636
;       FILING DATE:  16 JUL 1996
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/028,031
;       FILING DATE:  08 MAR 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Reiter, Stephen E.
;       REGISTRATION NUMBER:  31,192
;       REFERENCE/DOCKET NUMBER:  P41 9368
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619 546 4747
;       TELEFAX:  619 546 9392
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2450 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  both
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  267..1775
US 08 700 636 9


alignment scores:
            Quality:  104.50              Length:     374
              Ratio:  0.661                 Gaps:      20
 Percent Similarity:  42.246   Percent Identity:  22.460

alignment block:
 US 09 126 945B 2 x US 08 700 636 9/rev  ..

 Align seq 1/1  to reverse of: US 08 700 636 9  from: 1  to: 2450

        6 ProGlyLeuSerSerValSerProSerHisLeu..... 16
          ||||||||||||          ||||||:::|||
     2244 CCAGGCCTCTCTCTTCCCCTCCCTTCAGAGCTACCATTGTGTCAGGCCAG 2195

       17 .................LeuLeuProProAspThrValSerArgThrGlyL 28
                          |||||||||          :::|||   |||   |
     2194 CCCGGTGCGGTGTGAGCTCCTGCCCTTCCACCCCTTGTCCTATACCCTC 2145

       28 euGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSer 44
          ||          :::|||:::                 |||:::
     2144 TTCTCCCATGACTTGGAGGCAGC..............AGACAATCCTGT 2110

       45 ProSerPro.........ProAlaThrProGluGlnGly..... 54
          |||   |||          |||:::||||||:::   |||
     2109 CCCCTTCCTGGCCCCTTGCCCGTCACTCCTCAGGGTGGATGGCTCTGCCA 2060

       55 .............LeuSerAlaPhe.....................I 59
                      ||||||:::::                      :
     2059 TCTCTGTCTTCCACTTAGCAGCCTTGCTCCCCTCCTCAACACTATTGCCT 2010

       59 yrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAla 75
          ::   :::       |||          |||:::::|||||||:::::
     2009 GGGCTGGTACAAGAGATGGAGCAGCAAAACAAAGCAGTTGGTCCAGC 1960

       76 LysAlaProGlyAlaSerSer.........ArgGluGluProProGluGl 89
                 |||||||:::          |||:::
     1959 CTCCAGGTTGGAGCTGCTGTGTGTGGATGGAGGCTGGGGGAGGAGGGAC 1910

       89 uProGluGluCysPro......ValIleAspSerGluAlaProAlaGlyS 104
          |||:::::::::          :::      ::::::||||||
     1909 TCCAGATGAAGCAAAAGGATGGATCTGTGGGGGCAGCGCCCCT..... 1867

      104 erLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGlu 120
                ::::|||||:::      ::::::
     1866 ..........CTTCCTGGTAGCTCACTCGTCCATCCTCCACCCAACACT 1818

      121 GlnValGlnSerMetValValGly..........GluValLeuLysAs 133
          |||||||::   |||:::|||            |||:::   |||
     1827 ACTGTGCAGCAGAGGGTGGCAGGGTGAAAGAGCATGGAGATGAGGAAGGG 1778

      133 pIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspt 150
          :::   :::    |||   ::::::::::::          }
     1777 CCTCACTTGGAGCTGGGGGCTGAGTGGTCTGAGTGGAGGAAGGTGGTGGT 1728

      150 rpSerProSerAsnValGln.............LysTrpLeuLeuTrp 161
          ||      |||:::   :::                 :::|||
     1727 GGTGTAGTTCTGGAAGAGACGCTGCAGGAACATGCCGATGGTGC..... 1684

      162 ThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGluLe 178
          :::|||:::   |||             |||:::::   ::::::
     1683 CAAAGACACAGACAAAGACAAAGATCCAGAGGAAGAGGCGGTCGATCACC 1634

      178 uAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgS 195
                :::                |||:::::      |||   ||||
     1633 ATGGCGACGTACTTCCAGTCCTCACTCACGCTCTGGTCATCGTCCTCGCT 1584

      195 erProLeuGly.GlyAspValLeuHisAlaHisLeuAspIleTrpLysSe 211
          ||      |||   |||     ||||||
     1583 CCGCATGTGGTCTGCGATGAAGCGCACGCC..................... 1554

      211 rAlaAlaTrpMetLysGluArgThr.................SerProG 222
           :::      :::   |||   |||                 :::|||
     1553 .GTCCACCGCCTCCCGGAGGCCACAGCCACACGGCTCCCCTGACGCCCG 1505

      222 lyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSer 238
          |||||   |||:::|||       :::   |||
     1504 GGGCCGCCACTGGTGCAGGCTCAGCCCCGAAGGC................ 1470

      239 GluValAspSerSerCysSerGlyGlnProIleHisLeu........... 251
                              ||||||||||:::|||
     1469 ..................CCCGGCCAACCCCTGCACCGACGCGCGTTGA 1438

      252 ........................TrpGlnPheLeuLysGlu..... 257
                                  |||   |||   :::|||
     1437 CGAAGCACGTGCAGGAGTCGGCCCCTGGGGCTCGCGGAAGAAGAGGGCT 1388

      258 .................IeuLeuLeuLysProHisSerTyr..... 265
                             |||   :::|||:::::
     1387 CCAGCGCCCTCGCGCTCACGCTGCGCTCGCCGCAGGCGCAGGCGCTGACG 1338

      266 GlyArgPheIleArgTrpLeu......AsnLysGluLysGlyIlePheLy 280
          |||   ::::::   ||||||      :::::::::::::|||   :::
     1337 GGCGCAATGATGGCGTGGCTGCTGCATGAAGACGCAGCGCGGCAGCTTCT 1288

      280 sIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnA 297
          |||||||:::   :::   ::::||      |||:::|||   :::|
     1287 CCAGGAAGACGACCTTCACTCAGGGCGC...CATGGTGTGCGTGGTGGGC 1241

      297 rgProAlaMetAsnTyrAsp 304
          ||   :::::::::   :::
     1240 GAGCGGTGGTGCACGTTGAG 1221

seq_name: ;  US 08 700 636 9  from 1 to 2450

seq_documentation_block:
; Sequence 9, Application US/08467574
```

```
; Patent No. 6022704
;  GENERAL INFORMATION:
;     APPLICANT:  Elliot, Kathryn J.
;     APPLICANT:  Ellis, Steven B.
;     APPLICANT:  Harpold, Michael M.
;     TITLE OF INVENTION:  HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
;     TITLE OF INVENTION:  RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
;     NUMBER OF SEQUENCES:  12
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Brown, Martin, Haller & McClaim
;       STREET:  1660 Union Street
;       CITY:  San Diego
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  92101-2926
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/467,574
;       FILING DATE:  June 5, 1995
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/028,031
;       FILING DATE:  March 8, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seidman, Stephanie L.
;       REGISTRATION NUMBER:  33,779
;       REFERENCE/DOCKET NUMBER:  6362-9949
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-238-0999
;       TELEFAX:  619-238-0062
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2450 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  both
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  267..1775
US-08-467-574-9


alignment_scores:
            Quality:  104.50              Length:     474
              Ratio:  0.661                Gaps:      20
 Percent Similarity:  42.246   Percent Identity: 22.460

alignment_block:
 US-09-126-945b-2 x US-08-467-574-9/rev   ..

 Align seq 1/1  to reverse of: US-08-467-574-9  from: 1  to: 2450

       6 ProGlyLeuSerSerValSerProSerHisLeu..................  16
         |:|||||||||||          |||||||:|||
    2244 CCAGGCCTCTCTCTTCCCCTCCCTTCAGAGCTACCATTGTGTCAGGCCAG 2195

      17 ................LeuLeuProProAspThrValSerArgThrGlyL  28
                         |||||||||          :::|||   |||   |
    2194 CCCGGTGCGGTGTGAGCTCCTGCCCTTCCACCCCTTGTCCTATACCCCTC 2145

      28 euGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSer  44
         ||           :::|||:::             |||:::
    2144 TTCTCCCATGACTTGGAGGCAGC..............AGACAATCCTGT 2110

      45 ProSerPro.........ProAlaThrProGluGluGly...........  54
         |||   |||         |||:::|||||:::   |||
    2109 CCCCTTCCTGGCCCCTTGGCCGTCACTCCTCAGGGTGGATGGCTCTGCCA 2060

      55 ..............LeuSerAlaPhe.........................  69
                       |||||||:::::                       :
    2059 TCTCTGTCTTCCACTTAGCAGCCTTGCTCCCCTCCTCAACACTATTGCCT 2010

      69 yrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAla  75
         ::   :::       |||         |||:::::::|||||||||:::::
    2009 GGGCTGGTACAAGACATGGACAGCAGCCAAAACAAAGCAGTTGGTCCAGC 1960

      76 LysAlaProGlyAlaSerSer.........ArgGluGluProProGluGl  89
                 ||||||:::            |||:::
    1959 CTCCAGGTTGGAGCTGCTGTGTTGTGGATGGAGGCGTGGGGGAGGAGGGAC 1910

      89 uProGluGlnCysPro......ValIleAspSerGlnAlaProAlaGlyS 104
         |||:::::::::          :::       :::::::||||
    1909 TCCAGATGAAGCAAAAGGATGGATCTGTGGGGGCAGCGCCCCT....... 1867

     104 erLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGlu 120
                 :::|||||:::      ::::::
    1866 ..........CTCCTGGTAGCTCACTCGTCCACCCTCCACCCAACACT 1828

     121 GluValGlnSerMetValVal:ly              GluVal:rpTyrAs 133
         |||||||:::   |||:::|||                |||:::   |||
    1827 ACTGTGCAGCAGAGGGTGGCAGGGTGAAAGAGCATGGAGATGAGGAAGGG 1778

     133 pIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspL 150
         :::       :::       |||  ::::::::::::            |
    1777 CATCACTTGGAG...AAGCTGACTGTCTGAGTGGAGGAAGTTGGTGGT 1728

     150 rpSerProSerAsnValGln.............LysTrpLeuLeuTrp 161
         ||      |||:::  :::              :::|||
    1727 GGTGTAGTTCTGGAAGAGAGCTGCAGGAACATGCCGATGGTGC..... 1684

     162 ThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnIleLe 178
         :::|||:::  |||           |||::::::  ::::::
    1683 CAAAGACACAGACAAAGACAAAGATCCAGAGGAAGAGGCGGTCGATCACC 1634

     178 uAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgS 195
            :::               |||:::::::     |||   ||||
    1633 ATGGCGACGTACTTCCAGTCCTCACTCACGCTCTGGTCATCGTCCTCGCT 1584

     195 erProLeuGly.GlyAspValLeuHisAlaHisLeuAspIleTrpLysSe 211
         ||      |||   |||      ||||||
    1583 CCGCCATGTGGTCTGCCATGAAGCGCACGCC..................... 1554

     211 rAlaAlaTrpMetLysGluArgThr.................SerProG 222
          :::    :::  |||  |||                   :::|||||
    1553 .GTCCACCGCCTCCCGGAGGCCACAGCCACACGGCTCCCCTGACGCCCG 1505

     222 lyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSer 238
         |||||   |||:::|||      :::   |||
    1504 GGGCCGCCACTGGTGCAGGCTCAGCCCCGAAGGC.............. 1470

     239 GluValAspSerSerCysSerGlyGlnProIleHisLeu..........  251
                          |||.|||||:::|||
    1469 ................CCCGGCCAACCCCTGCACCGACGCGCGGGTTGA 1438

     252 ........................IrpGlnPheLeuLysGlu.......  257
                                 |||   |||   :::|||
    1437 CGAAGCACGTGCAGGAGTCGGCCCCTGGGGCTTCGCGGAAGAAGAGGGCT 1388

     258 ................LeuLeuLeuLysProHisSerTyr.........  265
                         |||   :::|||:::::
    1387 CCAGCGCCCTCGCGCTCACGCTGGCCTCGCCGCAGGCGCAGCGCTGACG 1338

     266 GlyArgPheIleArgIrpLeu......AsnLysGluLysGlyIlePheLy 280
         |||  ::::::  ||||||      :::::::::::|||   :::
    1337 GGCGCAATGATGCGCTGGCTGCTGCATGAAGAGCAGCGCGGGCAGCTTCT 1288

     280 sIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnA 297
         ||||||:::   :::   :::|||      |||:::|||   :::|
    1287 CCAGGAAGACGACCTTCACCCAGGGCGC...CATGGTGTGCGTGGTGGGC 1241
```

```
      297 rqProAlaMetAsnTyrAsp 304
          ||   :::::::::   :::
     1240 GAGCGGTGGTGCACGTTGAG 1221

seq name: /cms2_f/ptodata/1/iga/0A_COMB_seq:US-09-036-987A-1

seq documentation block:
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;     APPLICANT:  Baltz, Richard H.
;     APPLICANT:  Broughton, Mary C.
;     APPLICANT:  Crawford, Kathryn P.
;     APPLICANT:  Madduri, Krishnamurthy
;     APPLICANT:  Merlo, Donald J.
;     APPLICANT:  Treadway, Patti J.
;     APPLICANT:  Turner, Jan R.
;     APPLICANT:  Waldron, Clive
;     TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;     TITLE OF INVENTION:  Production
;     NUMBER OF SEQUENCES:  39
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Dow AgroSciences LLC Patent Department
;       STREET:  9330 Zionsville Road
;       CITY:  Indianapolis
;       STATE:  Indiana
;       COUNTRY:  USA
;       ZIP:  46268
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/036,987A
;       FILING DATE:  09 MAR 1998
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Stuart, Donald R
;       REGISTRATION NUMBER:  28,479
;       REFERENCE/DOCKET NUMBER:  50,608
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (317)337 4816
;       TELEFAX:  (317)337 4847
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  80161 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1


alignment scores:
            Quality:  103.00              Length:     274
              Ratio:  0.811                Gaps:      16
 Percent Similarity:  46.450   Percent Identity: 25.547

alignment block:
 US-09-126-945B-2 x US-09-036-987A-1  ..

 Align seq 1/1  for US-09-036-987A-1  from: 1  to: 80161

       7 GlyLeuSerSerValSerProSerHisLeuLeuLeuProProAspThrVa 23
         |||          ::::||   ::::::||::::||          :::::
   32477 GGTCGACCGGTGATCTCGGTGGACCGACTCGTATTGCGGTCGGCTCCGCC 32526

      24 lSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuLeuA 10
         :   |||||||||   ::::::   |||      |||      |||   |
   32527 CCGGCGGACCGGTGCCGCCGCGACAGGCCCATCAAGCTL..CGGTTGTACC 32573

      40 rqArqAspTrp.....SerProSerProProAla 49
```

```
         ||   :::|||                    |||   :::||||||:::
   32574 GGTTGAGCTGGCCAACGGTTCAACTGCCGACATCCGCTCAGCCACCGTCC 32623

      50 ThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLe 66
                 |||   |||         :::|||         :::::
   32624 TGCGTGCTTCTCGGCACCTCAGAA.....GTGTCCGCTGACATACAGGT 32667

      66 uTyrProGlu...AspSerSerTrpAlaAlaLysAlaProGlyAlaSer 82
         :||||||:::     |||      ||||||          ||||||
   32666 GTATCGGGACCTCCGACTCGTGACGGCTGCGTTGGATGCCGGTGCCC.... 32713

      82 erArqGlnGluProProGluGluProGluGlnCysProValIleAspSer 98
              |||||||||                      |||:::   :::
   32714 .....GAACCACCC.................GGCGTCGTCATCGCA 32757

      99 GlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGl 115
         :::|||   |||                    ||||||   |||   ::
   32738 CCCACGCCCCCGGC.................GGTGGACGAACACGCGA 32769

     115 u.............GluHisSerLeuGluGlnValGlnSerMetValV 127
         :              :::|||:::|||:::   ||||||:::   :::
   32770 TGTCCGGGACACGACTCGGCATGCACTCGACCTGGTACAAGGCTGGCTTT 32819

     127 alGlyGluValLeuLysAspIleGluThrAlaCysLysLeuAsnIle 143
            :::   |||:::::::              :::|||||   :::
   32820 CCGATCAGCGACTCAACGAA.........TCCCGATTGCTCCTGGTG 32857

     144 ThrAlaAspProMetAspTrpSerPro.....SerAsnValGl 156
         |||        :::     :::|||          :::::::::
   32858 ACACAGGGAGCAGTGGCCGTGGAGCCGGGCGAACCCGTGACCGATCTGGC 32907

     156 nLysTrpLeuLeuTrp.....ThrGluHisGlnT 166
          :::   ||||||                |||||||||
   32908 GCAGGCCGCGCTCTGGGCACTGCTGCGGTCGACGCAGACCGAACACCCTG 32957

     166 yrArq.....LeuProProMetGlyLysAlaPheGlnGlu 177
           |||              :::|||     ::::::   :::
   32958 ATCGCTTCGTCGCTCGTCGATGTCCTGACCCGCGCAACTCCTCCCCCG 33007

     178 LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArqGln.. 193
         |||   |||         |||:::      :::      :::|||:::
   33008 CTGCCGGGGGGTGCTGGCCTGCGGCGAACCTCAGCTCGCGTTGCGACGTGG 33057

     194 .......ArqSerPro.....LeuGlyGlyAspValLeuHisA 204
               :::::|||          ||||||:::|||   :::
   33058 CGGCGCTCATGCGCCCAGACTGGCTGGACTGGGCAGCGATGACGTC.... 33103

     204 IaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGlyArqThrSer 220
           |||   :::        ::::::::::||        |||   |||
   33104 .....CTGCCGTGCGCGGACGGCACCGGCGTGGCGATTGGAGGCACGCGC 33148

     221 ProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThr.. 236
         |||||:::::::          |||   :::::|||   :::   |||
   33149 CCGGGAAGCCTGGATGGGTTGGCATTGGTGGACGAACCGACGGCCACGGC 33198

     237 ..........AspSerGluVal 240
                 |||:::|||||
   33199 ACCGCTGGGTGACCGTGAGGTC 33220

seq name: /cms2_f/ptodata/1/iga/5B_COMB_seq:US-08-249-276-6

seq documentation block:
; Sequence 6, Application US/08249276
; Patent No. 5840836
;  GENERAL INFORMATION:
;     APPLICANT:  Robin, Daniel
;     TITLE OF INVENTION:  PANCREATIC ISLET CELL ANTIGENS
;     TITLE OF INVENTION:  OBTAINED BY MOLECULAR CLONING
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;       STREET:  660 White Plains Road
;       CITY:  Tarrytown
```

```
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10591 5144
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;       COMPUTER:  Apple Macintosh
;       OPERATING SYSTEM:  System 7.5
;       SOFTWARE:  WordPerfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/239,276
;       FILING DATE:  05-MAY-1994
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/872,646
;       FILING DATE:  08-JUN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/715,181
;       FILING DATE:  14-JUN-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/441,703
;       FILING DATE:  04-DEC-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/312,543
;       FILING DATE:  17-FEB-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kurt G. Briscoe
;       REGISTRATION NUMBER:  33,141
;       REFERENCE/DOCKET NUMBER:  MCI 251.4-KGB
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (914) 332-1700
;       TELEFAX:  (914) 332-1844
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3243 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-239-276-6


alignment_scores:
            Quality:  101.50              Length:     134
              Ratio:  1.336                 Gaps:       8
 Percent Similarity:  56.716   Percent Identity: 27.612

alignment_block:
 US-09-126-945B-2 x US-08-239-276-6/rev   ..

 Align seq 1/1  ** reverse of  US-08-239-276-6  from  1  to  3243

      229 IrpLysSerAlaAlaTrpMetLysGlyArgThrCysProGlyAlaHis... 224
          ||   :::|||:::      ::::::   |||||||||||||||
     1241 TGGGCCGCGGCTTGTACTTGTAGTCAGGATACTTCTCCAGGTGCTGCCGG 1192

      225 .....HisTyrCysAlaSerThrSerGluGluSer1 205
                          |||     |||::::::        ||||
     1191 CTCAGCCGCGCCTGTTCCTCATAGTAGGGCTGCTTCTCCTGGTTGGTCAT 1142

      205 rpIhrAspSerGluValAspSerSerCys.SerGlyGlnProIleHisLe 251
          ||||::::|||||:::::::::|||||| :::::      :::||||:
     1141 GGACTTCCAGCGAGATCCAAGGATCTTGCTGATGCTGGAGTTGTGCATGT 1092

      251 uTrpGlnPheLeuLysGlu.....LeuLeuLeuLysProHisSer1 265
          :|||::   |||::::::        |||:::|||   ||||||
     1091 CTGGCAAGGCTTGCAGGATCTTCCTCCGCTGCATCCTTGGCCCACACCATG 1042

      265 yrGly.....Arg.PheIleArgTrp...LeuAsnLysGluLysGlyIl 278
            |||          ::::::   |||   |||   ::::::::::|||
     1041 AAGGCGTTCATGGGCCTCTTGATGTGGCTGCTGTTCGGGACTCGGGCAA 992

      278 ePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgL 295
          ::::::      :::   :::|||:::::::      :::
      991 GTGGCGGGACGCCATCCATGTCGCAGCTCAGCATGGCTTGCTCCAGTGGCT 942
```

```
      295 ysAsnArgPro.....AlaMetAsnTyrAspLysLeu 305
          :::::|||                |||   :::   :::::|||
      941 GCACACAGCCGTCCTCCAGCGGCTGCTTGGCTGGGGATGAGTCCAGGCTG 892

      306 SerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
            ||||||::|||            |||||:::::::   :::|||
      891 ATGAGGTCCTTACGG.....AAGGGGGTGTTGGGGGAGCCA 856

seq 1 aa  , code 1 of 1: 1 to 3 MW seq no 08 468 579B-6

seq_documentation_block:
; Sequence 6, Application US/08468579B
; Patent No. 5981700
;  GENERAL INFORMATION:
;     APPLICANT:  Rubin, Daniel
;     TITLE OF INVENTION:  PANCREATIC ISLET CELL ANTIGENS
;     TITLE OF INVENTION:  OBTAINED BY MOLECULAR CLONING
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;       STREET:  660 White Plains Road
;       CITY:  Tarrytown
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10591-5144
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;       COMPUTER:  Apple Macintosh
;       OPERATING SYSTEM:  System 7.5
;       SOFTWARE:  WordPerfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/468,579B
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/239,276
;       FILING DATE:  05 MAY-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/872,646
;       FILING DATE:  08 JUN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/715,181
;       FILING DATE:  14-JUN-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/441,703
;       FILING DATE:  04-DEC-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/312,543
;       FILING DATE:  17-FEB-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kurt G. Briscoe
;       REGISTRATION NUMBER:  33,141
;       REFERENCE/DOCKET NUMBER:  MCI 251.5-KGB
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (914) 332-1700
;       TELEFAX:  (914) 332-1844
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3243 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-468-579B-6


alignment_scores:
            Quality:  101.50              Length:     134
              Ratio:  1.336                 Gaps:       8
 Percent Similarity:  56.716   Percent Identity: 27.612

alignment_block:
 US-09-126-945B-2 x US-08-468-579B-6/rev   ..
```

```
Align_seq_1/1  to reverse of: US-08-468-577B-6  from: 1  to: 4243

     209 TrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIle.. 224
         |||   ::|||:::       ::::::   |||||||||||||||
    1241 TGGGCCGCGGCTTGTACTTGTAGTCAGGATACTTCTCCAGGTGCTGCCGC 1192

     225 .....HisTyrCysAlaSerThrSerGluGluSerI 245
                  |||     |||::::::          ||||
    1191 CTCAGCCGCGCCTGTTCCTCATAGTAGGGCTGCTTCTCCTGGTTGGTCAT 1142

     246 rpThrAspSerGluValAspSerSerCys.SerGlyGlnProIleHisLe 251
         |||||::|||||||::::::::::|||||| :::::     :::|||::
    1141 GGACTTCCAGCGAGATCCAAGGATCTTGCTGATGCTGGAGTTGTGCATGT 1092

     251 uTrpGlnPheLeuLysGlu.........LeuLeuLeuLysProHisSerT 265
         :|||::   |||::::::          |||::|||   ||||||
    1091 CTGGGAAGGCTTGCAGGATCTTCCTCCGCTCATCCTTGGCCCACACCATG 1042

     265 yrGly.....Arg.PheIleArgTrp...LeuAsnLysGluLysGlyIl 278
           |||          :::::   |||   |||   ::::::::::|||
    1041 AAGGCGTTCATGGGCCTCTTGATGTGGCTGCTGTTTCGGGACTCGGGGAA 992

     278 ePheLysIleGluAspSerAlaGluValAlaArgLeuTrpGlyIleArgL 295
          ::::::      :::   :::|||::::::         :::
     991 GTGGCGGGAGCCATCCATGTCGCAGCTCAGCATGGCTTCCTCCAGTGGGT 942

     295 ysAsnArgPro.....AlaMetAsnTyrAspLysLeu 305
         :::::::|||              |||   :::   ::::::|||
     941 GCACACAGCCGTCCTCCAGCCGCTCCTTGGCTGGGGATGAGTCCAGGCTG 892

     306 SerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
           |||||||::|||            |||||||::::::   :::|||
     891 ATGAGGTCCTTACGG.....AAGGGGGTGTTGGGGGAGCCA 856

Seq_name: /cgn2_6/ptodata/1/ima/6A_COMB.seq:US-08-468-577B-6

Seq_documentation_block:
; Sequence 6, Application US/08468577B
; Patent No. 6001804
;  GENERAL INFORMATION:
;     APPLICANT:  Rabin, Daniel
;     TITLE OF INVENTION:  PANCREATIC ISLET CELL ANTIGENS
;     TITLE OF INVENTION:  OBTAINED BY MOLECULAR CLONING
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;       STREET:  660 White Plains Road
;       CITY:  Tarrytown
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10591-5144
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;       COMPUTER:  Apple Macintosh
;       OPERATING SYSTEM:  System 7.5
;       SOFTWARE:  WordPerfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/468,577B
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/239,276
;       FILING DATE:  05-MAY-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/872,646
;       FILING DATE:  08-JUN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/715,181
;       FILING DATE:  14-JUN-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/441,704
;       FILING DATE:  04-DEC-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/312,543
;       FILING DATE:  17-FEB-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kurt G. Briscoe
;       REGISTRATION NUMBER:  33,141
;       REFERENCE/DOCKET NUMBER:  MPI 251.9 KGB
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (914) 332-1700
;       TELEFAX:  (914) 332-1844
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4243 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-468-577B-6


alignment_scores:
            Quality:  101.50              Length:    134
              Ratio:  1.336                 Gaps:      8
 Percent Similarity:  56.716   Percent Identity: 27.612

alignment_block:
 US-09-126-945B-2 x US-08-468-577B-6_rc2  ..

 Align_seq_1/1  to reverse of: US-08-468-577B-6  from: 1  to: 4243

     209 TrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIle.. 224
         |||   ::|||:::       ::::::   |||||||||||||||
    1241 TGGGCCGCGGCTTGTACTTGTAGTCAGGATACTTCTCCAGGTGCTGCCGC 1192

     225 .....HisTyrCysAlaSerThrSerGluGluSerI 245
                  |||     |||::::::          ||||
    1191 CTCAGCCGCGCCTGTTCCTCATAGTAGGGCTGCTTCTCCTGGTTGGTCAT 1142

     246 rpThrAspSerGluValAspSerSerCys.SerGlyGlnProIleHisLe 251
         |||||::|||||||::::::::::|||||| :::::     :::|||::
    1141 GGACTTCCAGCGAGATCCAAGGATCTTGCTGATGCTGGAGTTGTGCATGT 1092

     251 uTrpGlnPheLeuLysGlu.........LeuLeuLeuLysProHisSerT 265
         :|||::   |||::::::          |||::|||   ||||||
    1091 CTGGGAAGGCTTGCAGGATCTTCCTCCGCTCATCCTTGGCCCACACCATG 1042

     265 yrGly.....Arg.PheIleArgTrp...LeuAsnLysGluLysGlyIl 278
           |||          :::::   |||   |||   ::::::::::|||
    1041 AAGGCGTTCATGGGCCTCTTGATGTGGCTGCTGTTTCGGGACTCGGGGAA 992

     278 ePheLysIleGluAspSerAlaGluValAlaArgLeuTrpGlyIleArgL 295
          ::::::      :::   :::|||::::::         :::
     991 GTGGCGGGAGCCATCCATGTCGCAGCTCAGCATGGCTTCCTCCAGTGGGT 942

     295 ysAsnArgPro.....AlaMetAsnTyrAspLysLeu 305
         :::::::|||              |||   :::   ::::::|||
     941 GCACACAGCCGTCCTCCAGCCGCTCCTTGGCTGGGGATGAGTCCAGGCTG 892

     306 SerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
           |||||||::|||            |||||||::::::   :::|||
     891 ATGAGGTCCTTACGG.....AAGGGGGTGTTGGGGGAGCCA 856

Seq_name: /cgn2_6/ptodata/1/ima/US_COMB.seq:US-08-347-594A-1

Seq_documentation_block:
; Sequence 1, Application US/08347594A
; Patent No. 5849546
;  GENERAL INFORMATION:
;     APPLICANT:  Gartinkel, Leonard
;     APPLICANT:  Richter, Lamar
;     TITLE OF INVENTION:  CLONING AND PRODUCTION OF HUMAN VON
;     TITLE OF INVENTION:  WILLEBRAND FACTOR GPIb BINDING DOMAIN POLYPEPTIDES AND
;     TITLE OF INVENTION:  METHODS OF USING SAME
;     NUMBER OF SEQUENCES:  4
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/347,594A
;      FILING DATE:  No. 5849536ember 30, 1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  36547-B2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6153 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..6153
US-08-347-594A-1


alignment_scores:
            Quality:   98.00              Length:    287
              Ratio:  0.824               Gaps:     11
 Percent Similarity:  41.463   Percent Identity: 22.396

alignment_block:
 US-09-126-945B-2 x US-08-347-594A-1   ..

 Align seq 1/1  to: US-08-347-594A-1  from: 1  to: 6153

       2 GlySerAlaSerProGlyLeuSerSerValSerPro.....SerHisLe 16
         |||   |||||||||       :::|||:::   |||       |||
    1628 GGCTGCGCATCTCCCAGAACTCGGGTCCGCCTGGCCGTGGTGCACTACCAC 1677

      16 uLeuLeuPro...ProAspIhrValSerArgThrGlyLeuGluLysAlaA 42
             |||   |||   :::   ||||||||||||   :::       :
    1678 GACGGCTCCCACGCCTACATCGGGCTCAAGGACCCGGAAGCGACCATCAGA 1727

      32 laAlaGlyAlaVal............GlyLeuGluArgArgAspTrpSer 44
         :::::|||||||:::           :::::::::       |||
    1728 GCTGCGGCGCATTGCCAGCCAGGTGAAGTATGCGGGCAGCCAGGTGG... 1774

      45 ProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSe 61
             |||||||||
    1775 ...CCTCCACCAGCGAGG.............................. 1789

      61 rTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaP 78
                               :::   ::::::   ::::::|||:::
    1790 ......................TCTTGAAATACACACTGTTCCAAATCTT 1817

      78 roGlyAlaSerSerArgGluGluProProGluGluProGluGlnCysPro 94
         :::   |||:::      :::||||||   ::::||      |||
    1818 CAGCAAGATCGACCGCCCTGAAGCCTCCCGCATCGCCC......TGCCCC 1861

      95 ValIleAspSerGluAlaProAlaGlySerLeuAspLeuValProGlyL 111
         ::::::::::|||   |||                  ||||||
    1862 TCATGGTCAGCCAGGAGCCCCAACCGA............TGTCCCGGAAT 1899

     111 yLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValG 128
         |||:::      :::::::::   ::::::   :::||||:::
    1900 TTTGTCCGCTACGTCCAGGGCCTGAAGAAGAAGAAGGTCATTGTGATCCC 1949

     128 lyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThr 144
             :::|||       :::   |||:::   :::       :::
    1950 GGTGGGCATTGGGCCCCATGCCAACCTCAAGCAGATCCGCCTCATCGAGA 1999

     145 AlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTr 161
         :::   |||:::         ||||||            |||:::   ||
    2000 AGCAGGCCCCTGAGAACAAGGCCTTCGTGCTGAGCAGTGTGGATGAGCTG 2049

     161 pThrGluHisGlnTyrArg.........................LeuP 169
         |:::::         |||                        |||
    2050 GAGCAGCAAAGGGACGAGATCGTTAGCTACCTCTGTGACCTTGCCCCTGA 2099

     169 roProMetGlyLysAlaPheGluLeuAlaGlyLysGluLeuCysAla 185
         |||:::                              |||:::|||   |||
    2100 ACCCCTCCTCCTACTCTGCCCCCCGACATCGCCACAAGTCACTCTTGCCC 2149

     186 MetSerGluGluGlnPheArgLeArgSerProLeuGlyGlyAspValIe 202
         ::::::      ||||||      :::|||   |||         ::
    2150 CGGGGCTCTTGGGGGTTTCGACCCTGGGGCCCAAGAGGAACTCCATGGTT 2199

     202 nHisAlaHisLeuAspIleTrpSer..................... 211
         :         :::   |||||:::
    2200 CTGGATGTGGCGTTCTTCCTGGAAGATCGGACAAAATTGGTGAAGCCGA 2249

     212 .....................................AlaAlaTrpMet 215
                                             :::::|||||
    2250 CTTCAACAGGAGCAAGGAGTTCATGGAGGAGGTGATTCAGCGGATGCAT 2299

     216 LysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThr..... 230
             |||||||::       :::          |||:::|||||
    2300 TGGGCCAGGACAGCA......TCCACGTCACGGTGCTGCAGTACTCCTAC 2343

     231 .............SerGluGluSerTrpThrAspSerGluValAspSer 243
                     |||       |||         |||         :::|
    2344 ATGGTGACCGTGGAGTACCCCTTCAGCGAGGACAACAGTCCAAAGGGACAT 2393

     243 erCysSerGly 246
         ||||||||||
    2394 CCTGCAGCGGG 2404

seq_name: /s_nc/_/pt_data/_/ca/_/A_C_MF_seq_US_08_347_594a_2

seq_documentation_block:
; Sequence 2, Application US/08464682
; Patent No. 6008193
;  GENERAL INFORMATION:
;    APPLICANT:  Leonard Gartinkel, et al.
;    TITLE OF INVENTION:  Cloning and Production of Human Von
;    TITLE OF INVENTION:  Willebrand Factor GPIb Binding Domain Polypeptides and
;    TITLE OF INVENTION:  Methods of Using Same
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  1185 Avenue of Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,682
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  White, John P.
;       REGISTRATION NUMBER:  28,678
;       REFERENCE/Docket NUMBER:  35437-B2 Y
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 278-0400
;       TELEFAX:  (212) 391 0525
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6153 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI SENSE:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..6153
US 08 463 682 2


alignment_scores:
            Quality:   98.00              Length:     287
              Ratio:   0.824               Gaps:      11
 Percent Similarity:  41.463   Percent Identity: 22.300

alignment block:
 US 09 126 945B 2 x US 08 463 682 2  ..

 Align seq 1/1  to: US 08 463 682 2  from: 1  to: 6153

        2 GlySerAlaSerProGlyLeuSerSerValSerPro.....SerHisLe 16
          |||   |||||||||       :::|||:::   |||       |||
     1628 GGCTGCGCATCTCCCAGAAGTGGTTCCGCGTGGCCGTGGTGGACTACCAC 1677

       16 uLeuLeuPro...ProAspThrValSerArgThrGlyLeuGluLysAlaA 32
                |||   |||   :::   ||||||||||||   :::       :
     1678 GACGGCTCCCACGCCTACATCGGGCTCAAGGACCGGAACGCGACCATCAGA 1727

       32 LaAlaGlyAlaVal...........GlyLeuGluArgArgAspTrpSer 44
          :::::|||||:::            :::::::::            |||
     1728 GCTGCGGCGCATTGCCAGCCAGGTGAAGTATGCGGGCAGCCAGGTGG... 1774

       45 ProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSe 61
                |||||||||
     1775 ...CCTCCACCAGCGAGC.............................. 1789

       61 rTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaP 78
                                  :::   ::::::   ::::::|||:::
     1790 ......................TCTTGAAATACACACTGTTCCAAATCTT 1817

       78 roGlyAlaSerSerArgGluGluProProGluGluProGluGlnCysPro 94
            :::   |||:::       :::||||||   :::|||       |||
     1818 CAGCAAGATCGACCGCCTGAACCCTCCCGCATCGCCC......TGCTCC 1861

       95 ValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyL 111
                :::::::::|||   |||                |||||||
     1862 TGATGGCCAGCCAGGAGCCCCAACCGA............TGTCCCGAAC 1899

      111 yLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValVal 128
            |||:::     :::::::::   ::::::   :::|||:::
     1900 TTTGTCCGCTACGTCCAGGGCCTGAAGAAGAAGAAGGTCATTGTGATCCC 1949

      128 lyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThr 144
             :::|||       :::   |||:::   :::       :::
     1950 GGTGGGCATTGGGCCCCATGCCAACCTCAAGCAGATCCGCCTCATCGAGA 1999

      145 AlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTr 161
          :::   |||:::         ||||||            |||:::   ||
     2000 AGCAGGCCCCTGAGAACAAGGCTTCGTGCTGAGCAGTGTGGATGACCTG 2049

      161 pThrGluHisGlnTyrArg.........................LeuP 169
          |||||||         |||                         |||

     2050 GAGCAGCAAAGGGACGAGATCGTTAGCTACCTCTGTGACCTTGCCCCTGA 2099

      169 roProMetGlyLysAlaProGlnValLeuAlaGlyLysGluLeuCysAla 185
            |||:::                                |||:::|||   |||
     2100 AGCCCCTCCTCCTACTCTGCCCCCCGACATGGCACAACTTACTCTGGCCC 2149

      186 MetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValIle 202
          ::::::       ||||||       :::|||   |||           ::
     2150 CGGGGCTCTTGGGGGTTTCGACCTTGGGGCCCAAGAGGAACTCCATGGTT 2199

      202 nHisAlaHisLeuAspIleTrpLysSer..................... 211
          :            :::   |||||||:::
     2200 CTGGATGTGGCGTTCGTCCTGGAAGGATCGGACAAAATTGGTGAAGCCGA 2249

      212 .....................................AlaAlaTrpMet 215
                                                :::::|||||
     2250 CTTCAACAGGAGCAAGGAGTTCATGGAGGAGGTGATTCAGCGGATGGATG 2299

      216 LysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThr..... 230
               |||||||:::       :::          |||:::|||||
     2300 TGGGCCAGGACAGCA......TCCACGTCACGGTGCTGCAGTACTCCTAC 2343

      231 .............SerGluGluSerTrpThrAspSerGluValAspSerS 243
                      |||       |||         |||         :::|
     2344 ATGGTGACCGTGGAGTACCCCTTCAGCGAGGCACAGTCCAAAGGGACAT 2393

      243 erCysSerGly 246
          ||||||||||||
     2394 CCTGCAGCGGG 2404

Seq name  no. 1 of Patent US 08/463,682, Seq. 2, Ver. 2, 684C 1

seq_documentation block:
; Sequence 1, Application US/08721684C
; Patent No. 5854016
;  GENERAL INFORMATION:
;    APPLICANT:  Keegan, Kathleen S.
;    TITLE OF INVENTION:  No. 5854016el CREBa Isoform
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive, 6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/721,684C
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams Jr., Joseph A.
;      REGISTRATION NUMBER:  38,659
;      REFERENCE/Docket NUMBER:  27869/33487
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474 6300
;      TELEFAX:  312 474 0448
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3190 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  304..1866
US 08 721-684C 1
```

```
alignment_scores:
            Quality:   97.00              Length:    299
              Ratio:  0.630                Gaps:     18
 Percent Similarity:  51.505   Percent Identity: 24.415

alignment_block:
 US-09-126-945B-2 x US-08-721-684C-1    ..

 Align seq 1/1  to: US-08-721-684C-1  from: 1  to: 3190

       2 GlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLe 18
         ||:::::::|||   ::|||   :::   ||||||::   ||||||
     132 GGACAAGGGTCGCGTAGCCTTCGGCAGGGCTCTCCCGAGTTCCTGCTC.. 179

      18 uProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGly. 64
           :::   ::|||   |||      :::          ::||||
     180 .....CAGTGCATAAGTTCCACGCCGCCACACGCCAAGTACACGGGGA 222

      65 ..AlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
         :::|||:::      |||          ...   ||||||||||||...
     223 GAAGCGTCTCACCGGCCCGCCGCGGCTCTGCGCGGTCCCCTCCTGCCTCA 272

      51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
         ::::::|||||   :::
     273 GCATCCTCGGGCCTGCGGGGC.......................... 293

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
                        |||   :::      |||||||::       |
     294 ..............   GCCCACCGCCATGGAGGTCCTGCA      G 338

      84 luGluProProGluGluProGluGlnCysProValIleAspSerGlnAla 100
         ||:::      |||:::|||         |||    :::      :::
     339 AGCGGGGAGCAGAGCGTCCTGCAGTGGGACCGCAAGCTGAGCGAGCTGTC 388

     101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGlnH 117
         ||| ||||::   :::            ::|||     :::    |
     389 AGAGCCCGGAGAGACTGAG.........GCCCTCATGTACCACACGC 426

     117 is.......SerLeuGluGlnValGluSerMetValValGlyGluVal 133
         ||          |||::::::   ::::::   |||:::|||||::
     427 ACTTCTCGGAGCTCCTACACGACGTTTCCCACAAGGTCCTGGTCACGCTC 476

     134 LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspPr 147
         |||::::||         :::   |||      ::::::   :::::::|
     477 CTGAGTGACCCTTTCCTCTCAGAGAACAGCGAGTCAATGGAGGTGGAGCC 526

     147 oMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluH 164
         |        ||||||:::              ||:::   :::||||
     527 ATCTCCAACATCACCAGCGCCT.........CTCATCCAGCCTGAAC 574

     164 isGlnTyrArgLeu.........ProProMetGlyLysAlaPheGlnGlu 177
         ||::::||   |||       |||       :::   |||   :::
     575 ACAGCTACTCTCTGAGCGAGGAGCCCCGGACTCACTCACCATTACCCAT 624

     178 LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGln.PheArgGlnA 194
         |||   ::::::      ::::::::::::::::: ::::||   :
     625 GCGCCTACCAGCGAC......AGCTTCAATGACGAGGAGGTCGAGAGTGA 668

     194 rqSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLys 210
         :::::          |||      |||::::::   |||::::::
     669 AAAAT.........GGTACCTGCCTACAGAGTTTCCTTCAGCTA...CCA 716

     211 SerAlaAlaTrpMetLysGluArgThrSer.ProGlyAlaIleHisTyrC 227
         |||       :::::::::|||:::||| ||||||::       ::::
     717 ICAAGAAAGAGCCAATCACAGAGGAGCAGCCCCCGGGACTTGTCCCTCCT 766

     227 ysAlaSer.Thr.........................SerGluGluSe 234
         ||   ||| |||                        :::||||||||
     767 GTCACTCTGACCATCACAGCCATTTCCACTCCTTTTCAAAAAGAAGAGTC 816

     234 rTrpThrAspSerGluVal......AspSerSerCysSerGlyGlnProI 249
```

```
         |      |||   ::::::      ||||||||||||:::
     777 CCCTCTGGATATGAATGCTGGGGGGGACTCCTCATGCCAG.......... 816

     249 leHisLeuTrpGlnPheLeuLysGluLeuLeuLysProHisSerTyr 265
                      ::::::   ::::::   |||:::|||||:::
     817 ..........ACGCTTATTCCTAAGATTAAGCTGGAGCCCCACGAAGTG 865

     266 GlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
         :::|||::::::::::      ||||||   ::::::
     866 GATCAGTTCTTAAACTTCTCCCCGAAACAAGCCTCCGTG 904

seq_name: [illegible] seq-US-09-005-970-1

seq_documentation_block:
; Sequence 1, Application US/09005970
; Patent No. 5959079
;  GENERAL INFORMATION:
;     APPLICANT:  Keegan, Kathleen S.
;     TITLE OF INVENTION:  No. 59590790l CREBa Isoform
;     NUMBER OF SEQUENCES:  10
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;       STREET:  233 South Wacker Drive, 6300 Sears Tower
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  USA
;       ZIP:  60606
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/005,970
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/721,684
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Williams Jr., Joseph A.
;       REGISTRATION NUMBER:  38,659
;       REFERENCE/DOCKET NUMBER:  27866/33487
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312-474-6300
;       TELEFAX:  312 474-0448
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3190 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  304..1866
US-09-005-970-1


alignment_scores:
            Quality:   97.00              Length:    299
              Ratio:  0.630                Gaps:     18
 Percent Similarity:  51.505   Percent Identity: 24.415

alignment_block:
 US-09-126-945B-2 x US-09-005-970-1    ..

 Align seq 1/1  to: US-09-005-970-1  from: 1  to: 3190

       2 GlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLe 18
         ||:::::::|||   ::|||   :::   ||||||::   ||||||
     132 GGACAAGGGTCGCGTAGCCTTCGGCAGGGCTCTCCCGAGTTCCTGCTC.. 179
```

```
 18 uProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGly. 34
            :::   :::|||  |||      :::          :::|||
180 ......CAGTGCATAAGTTCCACGCGGCACACGCCAAGTACACGGGA 222

 35 ..AlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
      :::|||:::      |||      :::   |||||||||||||:::
223 GAAGCGTCTCACCGGCCCGCGGCGGCTCTGCGGCGTCCCCTCCTGCCTCA 272

 51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
      ::::::|||||   :::
273 GCATCCTCGGGCCTGCGCGGC.......................... 298

 67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
                    |||   :::      |||||:::      |
294 ................GCCCACCGCCATGGAGGTGCTGGA.....G 338

 84 luGluProProGluGluProGluGlnCysProValIleAspSerGlnAla 100
    ||:::      |||:::|||         |||      :::      :::
339 AGCGGGGAGCAGAGCGTCCTGCAGTGGGACCGCAAGCTGACGGAGCTGCC 388

101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGlnL 117
      ||| |||:::  :::          :::|||      :::      |
389 AGAGCCCGGAGAGACTGAG.........GCCCTCATGTACCACACGC 426

117 is.........SerLeuGluGlnValGlnSerMetValValGlyGluVal 130
    ||          |||:::::::  ::::::  ||||:::|||:::::
427 ACTTCTCGGAGCTCCTAGACGAGTTTTCCCAGAACGTCCTGGGTCAGCTC 476

131 LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspPr 147
    |||:::|||         :::   |||      ::::::   :::::::||
477 CTCAGTGACCCTTTCCTCTCAGAAGAGCGAGTCAATGGACGTGGAGCC 526

147 oMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluE 164
    |         |||||:::                 |||:::   :::||||
527 ATCTCCAACATCACCAGGCCT.............CTCATCCAGGCTGAAC 574

164 rsGlnTyrArgLeu.........ProProMetGlyLysAlaPheGlnGlu 177
    ||:::|||   |||         |||         :::   |||   :::
575 ACAGCTACTCTCTGAGCGAGGAGCCCCGGACTCAGTCACCATTTACCCAT 624

178 LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGln.PheArgGlnA 194
      |||   ::::::      :::::::::::::|||:: :::|||   :
625 GCGGCTACCAGCGAC......AGCTTCAATGACGAGGAGGTGGAGAGTGA 668

194 rgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLys 210
    :::::         |||      ||:::::::   ||||::::::
669 AAAAT.........GGTACCTGTCTACAGAGTTTCCTTCAGCTA...CCA 716

211 SerAlaAlaTrpMetLysGluArgThrSer.ProGlyAlaIleHisTyrC 227
    |||         :::::::::|||:::||| ||||||:::      :::|
717 TCAAGAAAGAGCCAATCACAGAGGAGCAGCCCCCGGGACTTGTCCCTTCT 766

227 ysAlaSer.Thr.........................SerGluGluSe 234
    ||   ||| |||                       :::||||||||
767 CTCACTCTGACCATCACAGCCATTTCCACTCCTTTGAAAAAGAAGAGCC 816

234 rTrpThrAspSerGluVal......AspSerSerCysSerGlyGlnProI 249
    |      |||   ::::::      ||||||||||||||:::
817 CCCTCTGGATATGAATGCTGGGGGGACTCCTCATGCCAG.......... 866

249 leHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyr 265
                ::::::   ::::::   ||||:::||||||:::
867 ..........ACGCTTATTCCTAAGATTAAGCTGGAGCCCCACGAAGTG 916

266 GlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
      :::|||:::::::::      ||||||   ::::::
917 GATCAGTTCTTAAACTTCTCCCCGAAAGAAGCCTCCGTG 954

seq name: /cgn2_6/prodata/1/iex/08_0?MR_seq:US-09-407-715-1

seq documentation block:
; Sequence 1, Application US/09407715
; Patent No. 6248532

;   GENERAL INFORMATION:
;        APPLICANT: Keegan, Kathleen S.
;        TITLE OF INVENTION: No. 6248532c1 CREBa Isoform
;        NUMBER OF SEQUENCES: 10
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;             STREET: 233 South Wacker Drive, 6300 Sears Tower
;             CITY: Chicago
;             STATE: Illinois
;             COUNTRY: USA
;             ZIP: 60606
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/407,715
;             FILING DATE: 28-Sep 1999
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/005,970
;             FILING DATE: 12 JAN 1998
;             APPLICATION NUMBER: US/08/721,684
;             FILING DATE: 27-SEP 1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Williams Jr., Joseph A.
;             REGISTRATION NUMBER: 48,659
;             REFERENCE/DOCKET NUMBER: 27866/34497
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 312-474-6300
;             TELEFAX: 312 474 0448
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 4190 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        FEATURE:
;             NAME/KEY: CDS
;             LOCATION: 804..1866
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-407-715-1


alignment scores:
          Quality:   97.00          Length:    299
            Ratio:  0.630            Gaps:     18
 Percent Similarity:  51.505   Percent Identity: 24.415

alignment block:
 US-09-126-945B-2 x US-09-407-715-1  ..

 Align seq 1/1  to: US-09-407-715-1  from: 1  to: 4190

       2 GlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLe 18
         |||:::::::|||   :::|||   :::   |||||||:::   ||||||
     132 GGACAAGGGTCGCGTAGCCTTCGGCAGGGCTCTCCCGAGTTCCTGCTC.. 179

      18 uProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGly. 34
            :::   :::|||  |||      :::          :::|||
     180 ......CAGTGCATAAGTTCCACGCGGCACACGCCAAGTACACGGGA 222

      35 ..AlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
           :::|||:::      |||          :::   |||||||||||||:::
     223 GAACCGTCTCACCGGCCCGCGGCGGCTCTGCGGCGTCCCCTCCTGCCTCA 272

      51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
           ::::::|||||   :::
     273 GCATCCTCGGGCCTGCGCGGC.......................... 293

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
                         |||   :::      |||||:::      |
```

```
     294 ..................GCCCACCGCCATGGAGGTGCTGGA......G 318

      84 InGluProProGluGluProGluGlnCysProValIleAspSerGluAla 100
         !:::: ::!:::!!! ::!!! ::: ::: :::
     319 AGCGGGGAGCAGAGCGTCCTGCAGTCGGACCGCAAGCTGACGAGGTTTC 368

     101 ProAla.GlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluH 117
         !!! !!!:: ::: ::!!!! ::: !
     369 AGAGCCCGGAGAGACTGAG..........GCCCTCATGTACCACACGC 413

     117 is.....SerLeuGlyGluValGlnSerMetValValGlyGluVal 139
         !! !!!::::: :::::: !!!:::!!!::::::
     407 ACTTCTCGGAGCTCCTAGACGAGTTTTCCCAGAACGTCCTGGGTCAGCTC 456

     131 LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspPr 147
         !!!:::!!! ::: !!! :::::: :::::!!!
     457 CTGAGTGACCCTTTCCTCTCAGAGAAGAGCGAGTCAATGGAGGTGGAGCC 506

     147 oMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluH 164
         ! !!!!!!:: !!!:: ::: !!!
     507 ATCTCCAACATCACCAGCGCCT.....CTCATCCAGGCTGAAC 544

     164 isGluTyrArgLeu.....ProProMetGlyLysAlaPheGlyGln 177
         !:::!!. !!! !!! ::: !!! :::
     545 ACAGCTACTCTCTGAGCGAGGAGCCCCGGACTCAGTCACCATTTACCCAT 594

     178 LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGln.PheArgGlnA 194
         !!! :::::: :::::::::::!!:::: :::!!! :
     595 GCGGCTACCAGCGAC.....AGCTTCAATGACGAGGAGGTGGGAGAGTGA 638

     194 rgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLys 210
         ::::: !!! !!!::::: !!!:::::
     639 AAAAT.....GGTACCTGTCTACAGAGTTTCCTTCAGCTA...CCA 676

     211 SerAlaAlaTrpMetLysGluArgThrSer.ProGlyAlaIleHisTyrC 227
         !!! :::::::::::!:::!!! !!!!!!:: :::!
     677 TCAAGAAAGAGCCAATCACAGAGGAGCAGCCCCCGGGACTTGTCCCTTCT 726

     227 ysAlaSer.Thr.....SerGluGluSe 244
         !! !!! !!! :::!!!!!!!!
     727 GTCACTCTGACCATCACAGCCATTCCACTCCTTTGAAAAAGAAGAGTC 776

     244 rTrpThrAspSerGluVal.....AspSerSerCysSerGlyGlnProI 259
         ! !!! :::::: !!!!!!!!!!!!:::
     777 CCCTCTGGATATGAATGCTGGGGGGGACTCCTCATGCCAG..... 816

     259 leHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyr 265
                 :::::: :::::: !!!:::!!!!!!:::
     817 ..........ACGCTTATTCCTAAGATTAAGCTGGAGCCCCACGAAGTG 855

     266 GlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
         :::!!!!!!!!!!!. !!!!!! :::::
     856 GATCAGTTCTTAAACTTCTCCCCGAAAGAAGCCTCCGTG 894

seq_name: "/a1/g1/data1/1_1_5B_COMB.seq:US-08-982-956-1

seq_documentation_block:
; Sequence 1, Application US/08982956
; Patent No. 5861312
;  GENERAL INFORMATION:
;    APPLICANT:  Varshavsky, Alexander
;    APPLICANT:  Kwon, Yong Tae
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING MAMMALIAN UBR1
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  US
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/982,956
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  CIT-2001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (207) 363-0598
;      TELEFAX:  (207) 363-0528
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6395 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  115..5385
US-08-982-956-1


alignment_scores:
            Quality:   97.00              Length:    434
              Ratio:   0.591                Gaps:     21
 Percent Similarity:   67.788   Percent Identity: 18.433

alignment_block:
 US-09-126-945B-2 x US-08-982-956-1  ..

 Align seq 1/1  to: US-08-982-956-1  from: 1  to: 6395

      47 GlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGluGl 93
         !!! !!!!!! !!! !!!!!!::!!!!!! :::!!!:::
     146 GGCGCCGAGAGGAGGACGTCAGCCCGGAGCCTCCCCTGGCCCCGCAGCG 195

      93 nGlyLeuSer.....AlaPheTyrLeuS 61
         : !!! ::: :::!!!
     196 GCCGGCATCGTGGTGGGATCAGCAAGTTGATTTCTATACTGCTTTCTTAC 245

      61 erTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla 77
         :::::: :::!!! !!!!!! :::!!!
     246 ATCATTTGGCACAATTAGTGCCAGAA...ATTTATTTTGCTGAGATGGAC 282

      78 ProGlyAlaSerSerArgGluGluPro..... 86
         !!! :::::::::!!!!!!
     283 CCAGATTTGGAAAAGCAAGAAGAGAGTGTACAGATGTCAATACTCACTCC 332

      87 .....ProGluGluProGluGlnCysProValIleA 97
                 !!!:::!!!:: !!! ::::
     333 TTTGGAGTGGTACTTATTTGGAGAGGATCCGGATATTTGC.....TTAG 376

      97 spSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThr 113
         ::::: :::!!!:::::::::!!!
     377 AGAAATTAAAACACAGTGGAGCGTTCCAGTG..... 408

     114 LeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
                                                 !!!:::!!
     409 .....TGTGGGAAGGT 419

     130 lLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspP 147
         !:::!!!::: !!!!!!::: ::: ::: ::: !!!!
     420 TTTCAAAAGTGGAGAAACAACATATTCCTCTAGGGATTGTGCAATTGATC 469

     147 ro.....MetAspTrpSerProSerAsnValGlnLysTrp 158
         !! !!!!!! !!!:::!!!:::!!!
     470 CAACGTGTGTGCTCTGTATGGACTGCTTCCAAAGTAGTGTTCATAAA... 516
```

```
  159 LeuLeuTrpThrGluHisGlnTyrArgLeu...ProProMetGlyLysAl 174
                 :::|||:::|||:::::             |||   ::
  517 ...........AACCATCGTTACAAGATGCATACTTCTACTGGAGGGGC 554

  174 aPheGlnGluLeuAlaGlyLysGluLeu..................... 183
      :|||   :::   :::       |||
  555 CTTCTGTGACTGTGGAGACACAGAAGCTGGAAAACTGGCCCTTTTGTGC 604

  184 .....................................................Cys 184
                                                         |||
  605 TGGATCACGAGCCTGGAAGAGCAGGTACTACAAAAGAGAGCTTACATTGC 654

  185 AlaMetSerGluGluGlnPheArgGluArgSerProLeuGlyGlyAspVa 201
      ::::::|||||   :::   |||       :::       ::|||
  655 CCATTGAATGAAGAGGTGATTGCTCAAGCCAGGAGAATATTCCCTTCGGT 704

  201 HisHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluA 218
      |:::               |||   :::   :::   |||   :::|||:
  705 GATAAAATAC.........ATTGTAGAAATGACTATATGGGAAGAAGAAA 745

  218 rgThrSerProGlyAlaIle.....................HisTyr 226
      ::      |||      :::                     :::|||
  746 AGGAATTGCCTCCTGAACTGCAGATAAGGGAGAAAAATGAACGATACTAT 795

  227 CysAla..... 228
      ||:::
  796 TGTGTCTTTTCAACGATGAGCACCATCGTATGATCATGTGATCTACAG 845

  228 ..... 228

  846 TCTGCAGAGAGCTCTAGATTGCGAGCTTGCAGAGGCACAGCTGCACACGA 895

  229 .....................................................Ser 229
                                                         :::
  896 CTGCCATCGACAAAGAGGGTCGCCGGGCTGTCAAAGCAGGTGTGTATGCC 945

  230 ThrSerGluGluSerTrpThrAspSerGluValAspSer...SerCysSe 245
      |||   :::|||:::       |||   :::       |||   :::   ||
  946 ACTTGCCAGGAAGCAAAGGAGGATATAAAGAGTCACTCAGAGAACGTCTC 995

  245 rGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuGlySer 262
      |   :::|||:::|||:::   :::   |||   ::::::::::::
  996 TCAGCACCCCCTCCATGTG...GAAGTGCTGCACTCCGTGGTTATGGCTC 1042

  262 roHisSerTyrGly...ArgPheIleArgTrpLeuAsnLys..... 274
        ::::::::::::   ||:::      |||:::||||||
 1043 ACCAGAAATTCGCTCTGCGCCTTGGCTCCTGGATGAACAAAATTATGAGC 1092

  275 .....................................................Gl 275
                                                           ||
 1093 TATCAAGTGACTTTAGACAGATATTTGCCAGGCCTGCCTCGTAGAAGA 1142

  275 GLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArg.....L 290
      |   |||           |||:::   :::   ::::::|||      |
 1143 ACCTGGC.........ICTGAAAATCCCTGTCTTATAAGCAGACTAATGC 1184

  290 euTrpGlyIleArgLysAsnArgProAla..................... 299
      |||||      :::       :::   |||
 1184 TTTGGGATGCAAAACTTTATAAAGGTGCCCGTAAGATCCTTCATGAATTG 1233

  300 ........................MetAsnTyrAspLysLeu...SerAr 307
                              |||:::|||   ||||||   :::
 1234 ATCTTTAGTACTTTTTTTATGGAGATGGAATACAAAAAACTCTTTGCTAT 1283

  307 qSerIleArgGluTyrTyrLysLysGlyIleIleArgLysProAspIleS 324
      ::::::   :::||||||||||::       :::::|||      ||||
 1284 GGAATTTGTGAAGTATTATAAACAA......CTGCAGAAACAGTACATCA 1327

  324 er 324
      ||
 1328 GC 1329

seq name: Seq1 Pat Data_Lysis_6A.?? MB.seq.US-09-228-817-1
```

```
seq documentation block:
; Sequence 1, Application US/09228817
; Patent No. 6159742
;  GENERAL INFORMATION:
;     APPLICANT:  Varshavsky, Alexander
;     APPLICANT:  Kwon, Yong Tae
;     TITLE OF INVENTION:  NUCLEIC ACID ENCODING MAMMALIAN UBR1
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Kevin M. Farrell, P.C.
;       STREET:  P.O. Box 999
;       CITY:  York Harbor
;       STATE:  ME
;       COUNTRY:  US
;       ZIP:  03911
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/228,817
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Farrell, Kevin M.
;       REGISTRATION NUMBER:  35,505
;       REFERENCE/DOCKET NUMBER:  CIT 2001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (207) 363-0558
;       TELEFAX:  (207) 363 0528
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6495 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  115..5385
US-09-228-817-1


alignment_scores:
            Quality:   97.00              Length:    434
              Ratio:  0.591                Gaps:     21
 Percent Similarity:  47.788   Percent Identity: 18.433

alignment_block:
 US-09-126-945B-2 x US-09-228-817-1   ..

 Align seq 1/1  to: US-09-228-817-1  from: 1  to: 6495

      87 GlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGluGl 53
         |||   ||||||   |||   |||||||::||||||   :::|||:::::
     136 GGCGCCGAGAGGATGGACGTCAGCCCGGAGCCTCCCCTGGCCCCGCAGCG 185

      53 nGlyLeuSer.....................AlaPheTyrLeuS 61
         :   |||                      :::   :::|||
     186 GCCGGCATCGTGGTGGCATCAGCAAGTTGATTTCTATACTGCTTTCTTAC 235

      61 crTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla 77
         ::::::   :::|||   ||||||         :::|||
     236 ATCATTGGCACAATTAGTGCCAGAA...ATTTATTTGCTGAGATGGAC 282

      78 ProGlyAlaSerSerArgGluGluPro..................... 86
         |||      :::::::::||||||
     283 CCAGATTTGGAAAAGCAAGAAGAGAGTGTACAGATGTCAATACTCACTCC 332

      87 ...................ProGluGluProGluGluCysProValIleA 97
                             |||:::|||:::   |||      ::::
```

```
 333 TTTGGAGTGGTACTTATTTGGAGAGGATCCGGATATTTGC.....TTAC 376

  97 spSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThr 116
     :::::         :::|||||||||||||||
 377 AGAAATTAAAACACAGTGGAGCGTTCCAGTTG..................... 408

 114 LeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 133
                                                 |||:::||
 409 ..........................................TGTGGGAAGGT 419

 134 lLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspP 153
     |:::|||:::   |||||||:::  :::      :::   :::   ||||
 420 TTTCAAAAGTGGAGAAACAACATATTCCTGTAGGGATTGTGCAATTGATC 469

 147 ro..............MetAspTrpSerProSerAsnValGlnLysTrp 158
     ||              ||||||        ||||:::|||:::|||
 470 CAACGTGTGTGCTCTGTATGGACTGCTTCCAAAGTAGTGTTCATAAA... 515

 159 LeuLeuTrpThrGluHisGlnTyrArgLeu...ProProMetGlyLysAl 174
                 :::|||:::|||:::::             |||   ::
 517 ............AACCATCGTTACAAGATGCATACTTCTACTGGAGGGGG 564

 174 aPheGlnGluLeuAlaGlyLysGluLeu..................... 193
     :|||   :::   :::      |||
 565 CTTCTGTGACTGTGGAGACACAGAAGCGTGGAAAACTGGCCCTTTTGTG 614

 194 ..................................................Cys 194
                                                     |||
 615 TGGATCACGAGCCTGGAAGAGCAGGTACTACAAAAGAGAGCTTACATGC 664

 195 AlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspVa 214
     :::::|||||   :::   |||          :::      ::::|
 665 CCATTGAATGAACAGGTGATTGCTCAAGCCAGGAGAATATTCCCTTGGT 714

 215 lLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaIrpMetLysGluA 234
     |:::              |||   :::   :::   |||   ::::|||
 715 GATAAAATAC.........ATTGTAGAAATGACTATATGGGAAGAAGAAA 765

 235 rqfhrSerProGlyAlaIle.....................HisTyr 255
     ::      |||     :::                      :::|||
 766 AGGAATTGCCTCCTGAACTGCAGATAAGGGAGAAAAATGAACGATACTAT 815

 256 CysAla.......................................... 257
     ||||::
 816 TGTGTGCTTTTCAACGATGAGCACCATTCGTATGATCATGTGATCTACAC 865

 258 ..................................................... 258

 866 TCTGCAGAGAGCTCTAGATTGCGAGCTTGCAGAGGCACAGCTGCACACGA 915

 259 .....................................................Ser 259
                                                         :::
 916 CTGCCATCGACAAAGAGGGTCGCCGGCTGTCAAAGCAGGTGTATGCC 965

 260 ThrSerGlnGluSerTrpThrAspSerGluValAspSer...SerCysSe 279
     |||   :::|||:::      |||   :::    |||   :::   ||
 966 ACTTGCCAGGAAGCAAAGGAGGATATAAAGAGTCACTCAGAGAACGTCTC 1015

 280 rGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysP 299
     |   ::|||:::|||:::   :::   |||   ::::::::::::
1016 TCAGCACCCCCTCCATGTG...GAAGTGCTGCACTCCGTGGTTATGGCTC 1062

 300 roHisSerTyrGly...ArgPheIleArgTrpLeuAsnLys......... 314
       ::::::::::::   |||:::      |||:::|||||
1063 ACCAGAAATTCGCTCTGCGCCTTGGCTCCTGGATGAACAAAATTATCAGT 1112

 315 .....................................................Gl 315
                                                         ||
1113 TATTCAAGTGACTTTACACAGATATTTTGCCAGGCCTGCCTCGTAGAAGA 1162

 315 nLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArg.....L 339
     | |||          |||:::   :::   :::::|||      |
1163 ACCTGGC.........TCTGAAAATCCCTGTCTTATAAGCAGACTAATGC 1208

 290 euTrpGlyIleArgLysAsnArgProAla..................... 299
     |||||       :::     :::   |||
1184 TTTGGGATGCAAAACTTTATAAAGGTGCCCTAAGATCCTTCATGAATTG 1233

 300 .......................MetAsnTyrAspLysLeu...SerAr 307
                             |:|:::|||   ||||||   :::
1234 ATCTTTAGTAGTTTTTTTATGGAGATGGAATACAAAAAACTCTTTGCTAT 1283

 307 qSerIleArgGlnTyrIyrLysLysGlyIleIleArgLysProAspIleS 324
       ::::::   :::|||||||||||||:       ::::::|||      ||||
1284 GGAATTTGTGAAGTATTATAAACAA......CTGCAGAAAGAGTACATCA 1327

 324 or 324
     ||
1328 GC 1329
```


```
seq_name=" ... Sat Sep 18 17:08:993-118-1

seq_documentation_block:
; Sequence 1, Application US/08993118
; Patent No. 5997872
;  GENERAL INFORMATION:
;    APPLICANT:  LUCAS, Sophie;
;    APPLICANT:  DE SMET, Charles;
;    APPLICANT:  BOON-FALLEUR, Thierry
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
;    TITLE OF INVENTION:  REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Folie & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/993,118
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/845,528
;      FILING DATE:  April 25, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mary Anne Schofield
;      REGISTRATION NUMBER:  36,669
;      REFERENCE/DOCKET NUMBER:  LUD 5455
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688 9200
;      TELEFAX:  (212) 838 3884
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4031 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
US-08-993-118-1


alignment_scores:
            Quality:   96.50              Length:    201
              Ratio:  1.005                Gaps:      8
 Percent Similarity:  47.761   Percent Identity: 24.378

alignment_block:
 US-09-126-945B-2 x US-08-993-118-1   ..
```

```
Align_seq 1/1  to: US-08-993-118-1  from: 1  to: 4031

       3 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
         |||         |||   :::|||:::||||||   ::::::
    1697 AGCCCTCCTCAGGGGGAGGACTCCCTATCTCCTCACTACTTTCCTCAGAG 1746

      19 oPro.......... AspThrValSerArgThrGlyLeuGluLysAlaA 32
         |||          |||:::::::|||        :::   ::::::
    1747 CCCTCCCTCAGGGGGAGGACTCCCTGTCTCCTCACTACTTTCCTCAGAGCC 1796

      32 laAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProPro 48
            |||       :::|||:::   :::   :::       |||||||||
    1797 CTCAGGGGGAGGACTCCCTGTCTCCTCACTACTTTCCTCAGAGCCCTCCT 1846

      49 AlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMe 65
                |||:::::::::|||   :::|||:::             ::
    1847 ...CAGGGGGAGGACTCCATGTCTCCTCTCTACTTTCCTCAGAGTCCTCT 1893

      65 tLeuTyrProGlu.....................................A 70
         :          |||                                    :
    1894 TCAGGGGGAGGAATTCCAGTCTTCTCTCCAGAGCCCTGTGAGCATCTGCT 1943

      70 spSerSerTrpAlaAlaLysAlaProGlyAla......SerSerArgGlu 84
         :::|||||      ::::::   |||   :::       :::|||   :::
    1944 CCTCCTCCACTCCATCCAGTCTTCCCCAGAGTTTCCCTGAGAGTTCTCAG 1993

      85 GluProProGluGluProGluGlnCysProVal................. 95
         :::||||||||||   |||   |||   |||:::
    1994 AGTCCTCCTGAGGGGCCTGTCCAGTCTCCTCTCCATAGTCCTCAGAGCCC 2043

      96 ..........IleAspSerGlnAlaProAlaGlySerLeuAspLeuValP 109
                :::   |||||||:::|||       |||   :::   :::|
    2044 TCCTGAGGGGATCCACTCCCAATCTCCTCTCCAGAGTCCTGAGAGTGCTC 2093

     109 roGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMet 125
         ||   |||             |||   ||||||:::   :::|||
    2094 CTGAGGGG............GAGCATTCCCTGTCTCCTCTCCAA...... 2125

     126 ValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAs 142
            :::   :::         :::   |||   :::   :::   ||::
    2126 ...ATTCCTCAGAGTCCTCTTGAGGGAGAGGACTCCCTGTCTTCTCTCCA 2172

     142 nIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpL 159
         ::::       :::|||   :::|||:::   |||
    2173 TTTTCCTCAGAGTCCTCCTGAGTGGGAGGACTCC................ 2206

     159 euLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPhe 175
            |||          |||      ::::::|||||||:::|||:::   |||
    2207 ..CTCTCTCCTCTCCACTTTCCTCAGTTTCCTCCTCAGGGGGAGGACTTC 2254

     176 Gln 176
         |||
    2255 CAG 2257

Seq_name: /cqn2_6/prodata/1/inn/6A_COMB.seq:US-08-845-5280C-1

Seq_documentation_block:
; Sequence 1, Application US/08845528C
; Patent No. 6027924
;  GENERAL INFORMATION:
;     APPLICANT:  LUCAS, Sophie;
;     APPLICANT:  DE SMET, Charles;
;     APPLICANT:  BOON FALLEUR, Thierry
;     TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
;     TITLE OF INVENTION:  REJECTION ANTIGEN PRECURSOR MAGE-1 AND USES
;     TITLE OF INVENTION:  THEREOF
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lynch
;       STREET:  805 Third Avenue
;       CITY:  New York City
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10022
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  PC-DOS
;       SOFTWARE:  Wordperfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/845,528C
;       FILING DATE:  April 25, 1997
;       CLASSIFICATION:  4345
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Mary Anne Schofield
;       REGISTRATION NUMBER:  36,669
;       REFERENCE/DOCKET NUMBER:  LUD 5455
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688-9200
;       TELEFAX:  (212) 838-4884
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4031 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double stranded
;       TOPOLOGY:  linear
US-08-845-5280C-1


alignment scores:
             Quality:   96.50              Length:    201
               Ratio:   1.005               Gaps:      8
 Percent Similarity:  47.761   Percent Identity: 24.378

alignment block:
 US-09-126-945B-2 x US-08-845-5280C-1  ..

 Align_seq 1/1  to: US-08-845-5280C-1  from: 1  to: 4031

       3 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
         |||         |||   :::|||:::||||||   ::::::
    1697 AGCCCTCCTCAGGGGGAGGACTCCCTATCTCCTCACTACTTTCCTCAGAG 1746

      19 oPro..........AspThrValSerArgThrGlyLeuGluLysAlaA 32
         |||          |||:::::::|||        :::   ::::::
    1747 CCCTCCCTCAGGGGGAGGACTCCCTGTCTCCTCACTACTTTCCTCAGAGCC 1796

      32 laAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProPro 48
            |||       :::|||:::   :::   :::       |||||||||
    1797 CTCAGGGGGAGGACTCCCTGTCTCCTCACTACTTTCCTCAGAGCCCTCCT 1846

      49 AlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMe 65
                |||:::::::::|||   :::|||:::             ::
    1847 ...CAGGGGGAGGACTCCATGTCTCCTCTCTACTTTCCTCAGAGTCCTCT 1893

      65 tLeuTyrProGlu.....................................A 70
         :          |||                                    :
    1894 TCAGGGGGAGGAATTCCAGTCTTCTCTCCAGAGCCCTGTGAGCATCTGCT 1943

      70 spSerSerTrpAlaAlaLysAlaProGlyAla......SerSerArgGlu 84
         :::|||||      ::::::   |||   :::       :::|||   :::
    1944 CCTCCTCCACTCCATCCAGTCTTCCCCAGAGTTTCCCTGAGAGTTCTCAG 1993

      85 GluProProGluGluProGluGlnCysProVal................. 95
         :::||||||||||   |||   |||   |||:::
    1994 AGTCCTCCTGAGGGGCCTGTCCAGTCTCCTCTCCATAGTCCTCAGAGCCC 2043

      96 ..........IleAspSerGlnAlaProAlaGlySerLeuAspLeuValP 109
                :::   |||||||:::|||       |||   :::   :::|
    2044 TCCTGAGGGGATCCACTCCCAATCTCCTCTCCAGAGTCCTGAGAGTGCTC 2093

     109 roGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMet 125
         ||   |||             |||   ||||||:::   :::|||
    2094 CTGAGGGG............GAGCATTCCCTGTCTCCTCTCCAA...... 2125

     126 ValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAs 142
```

```
          :::   :::         :::   |||   :::   :::   |||::
    2126 ...ATTCCTCAGAGTCCTCTTCAGGGAGAGGACTCCCTGTCTTCTCTCCA 2172

     142 nIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpL 159
         ::::      :::|||   :::|||:::   |||
    2173 TTTCCTCAGAGTCCTCCTGAGTGGGAGGACTCC.............. 2206

     159 euLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPhe 175
           |||       |||       ::::::|||||:::|||:::   |||
    2207 ..CTCTCTCCTCTCCACTTTCCTCAGTTCCTCCTCAGGGGGAGGACTGC 2254

     176 Gln 176
         |||
    2255 CAG 2257

seq_name: /seq2_1/pub_data/726/US_08_993_118_9_COMB.seq; US-08-993-118-9

seq_documentation_block:
; Sequence 9, Application US/08993118
; Patent No. 5997872
;  GENERAL INFORMATION:
;    APPLICANT:  LUCAS, Sophie;
;    APPLICANT:  DE SMET, Charles;
;    APPLICANT:  BOON FALLEUR, Thierry
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
;    TITLE OF INVENTION:  REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 360 kb storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/993,118
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/845,528
;      FILING DATE:  April 25, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mary Anne Schofield
;      REGISTRATION NUMBER:  36,669
;      REFERENCE/DOCKET NUMBER:  LUD 5455
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4225 base pairs
;      TYPE:  nucleic acids
;      STRANDEDNESS:  double-stranded
;      TOPOLOGY:  linear
US-08-993-118-9


alignment_scores:
            Quality:   96.50              Length:    201
              Ratio:  1.005                Gaps:      8
 Percent Similarity:  47.761   Percent Identity: 24.378

alignment_block:
 US-09-126-945B-2 x US-08-993-118-9   ..

 Align seq 1/1  to: US-08-993-118-9  from: 1  to: 4225
```

```
        8 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
          |||         ||   :::|||:::||||||   ::::::
     1907 AGCCCTCCTCAGGGGGAGGACTCCCTATCTCCTCACTACTTTCCTCAGAG 1956

       19 oPro.............AspThrValSerArgThrGlyLeuGluLysAlaA 42
          |||              |||:::::::|||          :::   ::::::
     1957 CCCTCCTCAGGGGGAGGACTCCCTGTCTCCTCACTACTTTCCTCAGAGCC 2006

       42 laAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProPro 48
            |:|      :::|||:::   :::   :::       |||||||||
     2007 CTCAGGGGGAGGACTCCCTGTCTCCTCACTACTTTCCTCAGAGCCCTCCT 2056

       49 AlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMe 65
                |||:::::::::::|||   :::|||:::                ::
     2057 ...CAGGGGGAGGACTCCATGTCTCCTCTACTTTCCTCAGAGTCCTCT 2103

       65 tLeuTyrProGlu.....................................A 70
          :          ||:                                   :
     2104 TCAGGGGGAGGAATTCCAGTCTTCTCTCCAGAGCCCTGTGAGCATCTGCT 2153

       70 spSerSerTrpAlaAlaLysAlaProGlyAla......SerSerArgGlu 84
          :||||||       ::::::   |||   :::       :::|||   :::
     2154 CCTCCTCCACTCCATCCAGTCTTCCCCAGAGTTTCCCTGAGAGTTCTCAG 2203

       85 GluProProGlnGlnProGlnGlnCysProVal.................. 95
          :::|||||||||   |||   |||   |||:::
     2204 AGTCCTCCTGAGGGGCCTGTCCAGTCTCCTCTCCATAGTCCTCAGAGCCC 2253

       96 ..........IleAspSerGlnAlaProAlaGlySerLeuAspLeuVal 109
                    :::   |||||||:::|||       |||   :::   :::|
     2254 TCCTGAGGGATGTATCCCAATCTCCTCTCCAGAGTCCTGAGAGTGTCTC 2303

      109 roGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMet 125
          ||  |||             |||   |||||||:::   :::|||
     2304 CTGAGGGG............GAGGATTCCCTGTCTCCTCTCCAA...... 2345

      126 ValValGlyGlnValLeuLysAspIleGluThrAlaCysLysLeuLeuAs 142
             :::   :::         :::   |||   :::   :::   |||::
     2346 ...ATTCCTCAGAGTCCTCTTGAGGGAGAGGACTCCCTGTCTTCTCTCCA 2382

      142 nIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpL 159
          ::::      :::|||   :::|||:::   |||
     2383 TTTCCTCAGAGTCCTCCTGAGTGGGAGGACTCC.............. 2416

      159 euLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPhe 175
            |||       |||       ::::::|||||:::|||:::   |||
     2417 ..CTCTCTCCTCTCCACTTTCCTCAGTTCCTCCTCAGGGGGAGGACTGC 2466

      176 Gln 176
          |||
     2465 CAG 2467

seq_name: /seq2_1/pub_data/726/US_08_845_528C_9_COMB.seq; US-08-845-528C-9

seq_documentation_block:
; Sequence 9, Application US/08845528C
; Patent No. 6027924
;  GENERAL INFORMATION:
;    APPLICANT:  LUCAS, Sophie;
;    APPLICANT:  DE SMET, Charles;
;    APPLICANT:  BOON-FALLEUR, Thierry
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
;    TITLE OF INVENTION:  REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Diskette, 5.5 inch, 360 kb storage
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  PC DOS
;       SOFTWARE:  WordPerfect
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/845,528C
;       FILING DATE:  April 25, 1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Mary Anne Schofield
;       REGISTRATION NUMBER:  36,669
;       REFERENCE/DOCKET NUMBER:  LUD 5455
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 688 9200
;       TELEFAX:  (212) 838 3884
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4225 base pairs
;       TYPE:  nucleic acids
;       STRANDEDNESS:  double stranded
;       TOPOLOGY:  linear
US 08 845 528C 9


alignment scores:
            Quality:   96.50              Length:    201
              Ratio:  1.005                Gaps:      8
 Percent Similarity:  47.761   Percent Identity: 24.378

alignment block:
 US 09 126 945B 2 x US 08 845 528C 9   ..

 Align seq 1/1  to: US 08 845 528C 9  from: 1  to: 4225

        3 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
          |||          |||   :::|||:::||||||   ::::::
     1907 AGCCCTCCTCAGGGGCAGGACTCCCTATCTCCTCACTACTTTCCTCAGAG 1956

       19 oPro.............AspThrValSerArgThrGlyLeuGluLysAlaA 32
            |||           |||:::::::|||         :::   ::::::
     1957 CCCTCCTCAGGGGGAGGACTCCCTGTCTCCTCACTACTTTCCTCAGAGCC 2006

       32 laAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProPro 48
             |||      :::|||:::   :::   :::      |||||||||
     2007 CTCAGGGGGAGGACTCCCTGTCTCCTCACTACTTTCCTCAGAGCCCTCCT 2056

       49 AlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMe 65
                |||:::::::::|||   :::|||:::          ::
     2057 ...CAGGGGGAGGACTCCATGTCTCCTCTCTACTTTCCTCAGAGTCCTCT 2103

       65 tLeuTyrProGlu.....................................A 70
          :   |||                                         :
     2104 TCAGGGGCAGGAATTCCAGTCTTCTCTCCCAGAGCCCTGTGAGCATCTGCT 2153

       70 spSerSerTrpAlaAlaLysAlaProGlyAla......SerSerArgGlu 84
          ::|||||      ::::::   |||   :::       :::|||   :::
     2154 CCTCCTCCACTCCATCCAGTCTTCCCCAGAGTTTCCCTGAGAGTTCTCAG 2203

       85 GluProProGluGluProGluGlnCysProVal................. 95
          :::||||||||||   |||   |||   |||:::
     2204 AGTCCTCCTGAGGGGCCTGTCCAGTCTCCTCTCCATAGTCCTCAGAGCCC 2253

       96 ..........IleAspSerGluAlaProAlaGlySerLeuAspLeuVal 109
                    :::   |||||||:::|||      |||   :::   :::|
     2254 TCCTGAGGGATGCACTCCCAATCTCCTCTCTCAGAGTCCTGAGAGTGCTC 2303

      109 roGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMet 125
          ||   |||             |||   |||||||:::   :::|||
     2304 CTGAGGGG............CAGGATTCCCTGTCTCCTCTCCAA...... 2345

      126 ValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAs 142
            :::   :::         :::   |||   :::   :::   ||:::
     2346 ...ATTCCTCAGAGTCCTCTTGAGGGAGAGGACTCCCTGTCTTCTCTCCA 2394

      142 nIleThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrpL 159
          ::::      :::|||  :::||||::   |||
     2395 TTTTCCTCAGAGTCCTCCTGAGTGGGAGGACTCC.................. 2436

      159 euLeuTrpThrGluHisGluTyrArgLeuProProMetGlyLysAlaPhe 175
            |||         |||      ::::::||||||::|||:::   |||
     2437 ..CTCTCTCCTCTCCACTTTCCTCAGTTTCCTCCTCAGGGGGAGGACTTC 2464

      176 Gln 176
          |||
     2465 CAG 2467

Seq name: /tmp/ft2_31_pid_/t_31_t.tmp/68_COMB.seq:US 09 305 640 1

Seq documentation block:
; Sequence 1, Application US/09305640B
; Patent No. 6255468
; GENERAL INFORMATION:
;  APPLICANT: SmithKline Beecham plc
;  TITLE OF INVENTION: No. 6255468el Compounds
;  FILE REFERENCE: GP30124
;  CURRENT APPLICATION NUMBER: US/09/305,640B
;  CURRENT FILING DATE: 1999-05-05
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2560
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US 09-305-640 1


alignment scores:
            Quality:   96.00              Length:    498
              Ratio:  0.508                Gaps:      22
 Percent Similarity:  37.952   Percent Identity: 19.880

alignment_block:
 US 09-126-945B-2 x US 09-305-640 1   ..

 Align seq 1/1  to: US-09-305-640 1  from: 1  to: 2560

        4 AlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProPr 20
          |||:::|||:::   ::::::   ||||||                 |||
      635 GCCACGCCATCCCCGACGACAAGCTCACCTATGGCACCATCGCGGCCA.. 682

       20 oAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAla.... 35
             ::::::|||:::|||          ::::::|||   ||||||
      683 ....TCGGCGAGCAAAACG.........AGGAGCGCCTACGGCGCCTGCT 719

       36 ..............ValGlyLeuGluArgArg................... 41
                       |||||||||       |||
      720 GGCGGCGCCCGGGGCTGGGCCTGGCGGCGCGGCCCAGCGCAAGGTGCGCG 769

       42 ..........................................AspTrpSerPr 45
                                                   |||||||:::
      770 CCTTCCTTCCGCTCGTGCCTCGACATGCGCGAGATCGAGCGACTGGGCCCG 819

       45 oSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerT 62
          :::|||          ::::::      :::|||
      820 CGACCCATCCTAGACGTCATCGACGACTGCGGGGCT.............. 856

       62 yrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaPro 78
                                 ::::::||||||||||:::
      857 ........................GGGACCTGGGCGGCGCGGAGGAGC 880

       79 GlyAlaSerSerArgGluGlu.............................. 85
          :::|||||||:::::
      881 GTCCGGGGTCGCGGCGCGATGGGACCTCAACGGCTGCTGTACAAGGCG 930

       86 ................ProProGluGluProGluGlnCysProValIleA 97
                          ||||||::::::   ::::::      :::
```

```
 931 CAGGGCGTGTACAGCGCCGCCGCGCTCTTCTCGCTCACGGTCAGCCTGGA 980

  97 spSerGlnAlaProAlaGlySerLeuAspLeuVal.ProGly.GlyLeuT 113
       :::   ::::::   ::::::   :::   ::: ::::::  ::::::
 981 CGACAGGAACTCCTCGCGCTACGTCATCCGCATTGACCAGGATGGGCTCA 1030

 113 hrLeuGluGluHisSerLeu..............................  119
     :::::   ::::::::::::
1031 CCCTGCCAGAGAGGACCCTGTACCTCGCTCAGGATGAGGACAGTGACAAG 1080

 120 ..............................GluGlnValGlnSerMetValVa 127
                                  ::::::::   ::::::::
1081 ATCCTGGCAGCATACAGGGTGTTCATGGAGCGAGTGCTCAGCCTCCTGGG 1130

 127 lGlyGlnValLeuLysAsp..............IleGluThrAlaCysL 139
       ::::::::::::::::::            :::::::::       :
1131 TGCAGACGCTGTGGAACAGAAGGCCCAAGAGATCCTGCAAGTGGAGCAGC 1180

 139 ysLeuLeuAsnIleThrAlaAspProMetAsp.................. 149
     :::::   ::::::::::::::::       :::
1181 AGCTGGCCAACATCACTGTGTCAGAGTATGACGACCTACGGCGAGATGTC 1230

 150 ..........................................IrpSerPr 152
                                              :::::
1231 AGCTCCATGTACAACAAGGTGACGCTGGGGCAGCTGCAGAAGATCACCCC 1280

 152 oSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlaTyrArqLeuP 199
     |          ::::::::::::
1281 CCACTTGCGGTGGAAGTGGCTGCTA.......................... 1305

 169 roProMetGlyIyrAlaPheGlnGln   ....LeuAlaGlyIysGlnLeu 193
                :::   ::::::::                ::::::::
1306 ........GACCAGATCTTCCAGGAGGACTTCTCAGAGGAAGAGGAGGTG 1347

 184 CysAlaMetSerGluGluGlnPheArqGlnArqSerProLeu........ 197
           ::::::   :::   ::::::::   :::   :::
1348 GTGCTGCTGGCGACAGACTACATGCAGCAGGTGTCGCAGCTCATCCGCTC 1397

 198 ....GlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaA 218
                 :::::::::   :::::   :::::::::   ::::
1398 CACACCCCACCGGGTCCTGCACAACTACCTG...GTCTGGCGCGTGGTGC 1444

 219 laTrpMetLysGluArqThrSerPro.........GlyAlaIleHisTyr 229
     ::   ::::::::::::   ::::::         ::::::::::
1445 TGGTGCTGAGTGAACACCTGTCCCCGCCATTCCGTGAGGCACTGCACGAG 1494

 227 CysAlaSerThrSerGluGluSerTrpThrAspSerGluValAspSerSe 248
       ::::::       :::   :::          :::::::::
1495 CTGGCACAGGAGATGGAGGGCAGCGACAAGCCACAGGAGCTGGCCGGGT 1544

 248 rCysSerGlyGlnPro..................................  248
     :::   ::::::
1545 CTGCTTGCGCCAGGCCAATCGCCACTTTGGCATGGCGCTTGGCGCCCTCT 1594

 249 ..IleHis.....................................  250
       :::::
1595 ITGTACATGAGCACTTCTCAGCTGCCAGCAAAGCCAAGGTGCAGCAGCTA 1644

 251 .....................LeuTrpGlnPheLeuLysGluLeu..... 258
                         :::   :::   ::::::::::::
1645 GTGGAAGACATCAAGTACATCCTGGGCCAGCGCCTGGAGGAGCTGGACTG 1694

 258 ..................................................  258

1695 GATGGACGCCGAGACCACGGCTGCTGCTCGGGCCAAGTTCCACTACATGA 1744

 259 ........................LeuLeuLysProHisSerTyrGly 266
                             ::::::::::::   :::
1745 TGGTGATGGTCGGCTACCCGGACTTCCTGCTGAAACCCGATGCTGTGGAC 1794

 267 ArqPheIleArqTrpLeuAsnLysGluLysGlyIlePheLys...IleGl 282
     :::     ::::::       ::::::      ::::::   :::
1795 AAGCAGTATCAGTTTGAGGTCCATGAGAAGACCTACTTCAAGAACATCTT 1844

 282 uAspSerAlaGlnValAlaArqLeuTrpGlyIleArqLysAsnArq.... 297
       :::::::   :::   :::          ::::::::::::   :::
1845 GAACAGCATCCGCTTCAGCATCCAGCTCTCAGTTAAGAAGATTCGGCAGG 1894

 298 ..............................ProAlaMetAsnTyrAsp 308
                                    ::::::::::
1895 AGGTGGACAAGTCCACGTGGCTGCTCCCCCCACAGGCGCTCAATGCCTAC 1944

 304 LysLeuSerArqSerIleArqGlnTyrTyrLysLysGlyIleIleArq.. 319
           :::   ::::::   ::::::   :::       ::::::::::
1945 TATCTACCCAACAAGAACCAGATGGTGTTCCCCGGGGGCATCCTGCAGCC 1994

 320 ..........LysProAspIleSerGlnArqLeuValTyr 329
                :::::::::   :::   :::   :::
1995 CACCCTGTACGACCCTGACTTCCCACAGTCTCTCAACTAC 2034
```

seq_data: Patent US 08340428B Seq ID NO 1 5191 bp

```
seq_documentation_block:
; Sequence 1, Application US/08340428B
; Patent No. 5648465
;  GENERAL INFORMATION:
;    APPLICANT:  MARGOLIS, Richard U.
;    APPLICANT:  RAUCH, Uwe
;    APPLICANT:  MARGOLIS, Renee K.
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES FOR A
;    TITLE OF INVENTION:  NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Browdy and Neimark
;      STREET:  419 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/340,428B
;      FILING DATE:  14 No. 5648465ember 1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/922,911
;      FILING DATE:  03 August 1992
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Browdy, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  Margolis 1A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737 3528
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5191 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  77..3847
US 08 340 428B-1


alignment_scores:
            Quality:   95.00              Length:    401
              Ratio:   0.565               Gaps:     27
 Percent Similarity:  41.895   Percent Identity: 24.439
```

```
Alignment block:
 US-09-126-945B-2 x US-08-340-428B-1   ..

 Align seq 1/1  to: US-08-340-428B-1  from: 1  to: 5191

       5 SerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProProAs 21
         ||||||::|||      :::||||||||   :::         ||||:
    1766 TCCCCGTCGCTG......ATTTCACCCAGTACTGTC.......CCGAG 1800

      21 pThrValSerArgPheGlyLeuGluLysAlaAlaAlaGlyAlaValGly. 37
         :|||  |||       ||||||:::  :::|||      |||  |||
    1801 CACTGACAGTACTCCTGGCCTGAAGCCAGGGGCAGATGAGGCCCCTGCTG 1850

      38 ..........LeuGluArgArgAspTrpSerProSerProProAlaThr 50
                  :||||||||      |||  ||||||  ||||||||:::
    1851 TGAAGTCAGCCATCCACCACCCACCCTGGTTGCCCTCAGAACCCGCTGTC 1900

      51 ............ProGluGlnGlyLeuSerAla................. 57
                    |||:::::::::|||||||||
    1901 CCATCCTCCATTCCCTCAGAGGCCCTAAGTGCTGTCTCCCTACAGGCATC 1950

      57 ..................................................... 57

    1951 CCCTGGTGATGGCTCTCCAGACTTCCCCATTGTAGCCATGCTTCGAGCCC 2000

      58 .....PheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSer... 71
           :::::|||       :::   |||  ||:::   |||
    2001 CCAAACTGTGGCTTCTGCCACACTCTACACTCGTCCCGAATGTGTCCCCA 2050

      72 .........SerTrpAlaAlaLysAlaProGlyAlaSerSerArgGlu.. 84
                |||  ||:::      |||::::::      ::::||
    2051 ATCCCACTCTCCCCAGCTTCTCCACTCCCCTCCTCGGTCCCAGAAGAACA 2100

      85 ............................GluProProGluGluProGluG 92
                                    |||  ||||||  |||   |
    2101 GGCTGTCAGACCTGTCAGCTTTGGAGCAGAAGACCCGAGACCCCATTTC 2150

      92 In.....................................CysProValIle 95
         ||                                    |||
    2151 AGACCACCATGGCTGCCCCAGGTGAAGCCAGCCACGGATCCCCTGAGGCA 2200

      97 AspSer.................................................. 98
         ||||||
    2201 GACTCCATAGAAATCGAGGGGATCAGCTCCATGCAGGCTACAAAGCACCC 2250

      99 ....GlnAlaProAlaGlySerLeuAspLeu..................... 107
           ::::::|||   :::||||||||||
    2251 CATCTCTGGCCCATGGGCTTCTTTGGACTCCAGTAATGTGACACTGAATC 2300

     108 ..ValPro.........GlyGlyLeuThrLeuGluGluHisSerLeuGlu 128
           ||||||         |||  |||      ||||::      ||||::
    2301 CTGTCCCTTCTGATGCTGGCATCCTAGGGACTGAGTCTGGGGTCTTGGAC 2350

     129 GluValGlnSerMetValVal...GlyGluValLeuLysAspIleGluTh 146
               |||   :::      |||::::::      ||||::
    2351 TTACCAGGGAGTCCCACATCAGACGGACAGGCCACTGTGGACATGCTGCT 2400

     146 rAlaCysLysLeu.................LeuAsn.IleThrAlaAsp 160
          |||      |||               ||||:: :::::   :::
    2401 GGCCACCTGGCTACCACTGCCTGGCCACGGACTGGACACTGGCTCCCAGT 2450

     147 ProMetAspTrpSerProSer.................AsnValGlnLy 157
         |||      |||:::|||                  ||||:::|||
    2451 CCACACCCATGGAAGCCCATGGAGTAACCATGAGTGTGGAACCTACAG.. 2498

     157 stripLeuLeuTrpThrGluHisGlnIyrArgLeuProProMetGlyLys. 173
          |||   ||||||   |||             :::|||||      :::
    2499 .TGG...CTTTGGAAGGAG............GTGCCACCAAAGACCCAAA 2542

     174 AlaPheGlnGluLeuAlaGlyLysGluLeuCys.............AlaMe 186
         :::   ::::::   :::|||   :::   |||             :::::
    2543 GGAGGCCACCATGGATGTGGTCCCCAGCACTGTTGATGCCACTTCGGGGT 2592

     186 tSerGlnGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeu. 202
         :      ::::::|||::          |||   |||   |||
    2593 CTGAACCCAAAAGTTCCATTTCTAGCAGCCCATGTGGTTGTGACTGCAGCT 2642

     203 ...........HisAlaHisLeuAspIleTrpLys.SerAlaAlaTrpMe 215
                   |||:::|||   |||   :::|||  |||   :::   ::
    2643 GGGGACCAGGGCACACCCACACTGACCCCTACAAGCTCTGAAGGTCAGGT 2692

     215 tLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerG 232
         :     :::   |||   ||:::::          :::|||      :
    2693 GGTGGCCCAGGAGTCACTGGGAACCCTC.........ACCAGTCTGCCTT 2723

     232 IuGluSerTrpThrAsp..............SerGluValAspSerSer 243
         :::::   ||:::::::              :::||||||    |||
    2724 CTCATCCCTGGTCATCTCTGGCCTCCAGCATGGACGAAGTGGGCCTCGTT 2773

     244 CysSerGlyGlaProIleHisLeuTrpGlnPheLeuLysGluLeuLeuL 260
           ||||||:::|||   :::|||||||:::::   :::   ||||:: |
    2774 TCCTCAGGAGAACCCACAAGGTTGTGGGACATCCCCAGCACTCTGATACC 2823

     260 euLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLys 276
         ||   |||   :::                  ||:::|||
    2824 TGTGTCCTGGGGCT.................TGGATGAAT......... 2846

     277 GlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrp 291
                                    |||:::   ||||||||||
    2847 ..........................CAGACCTGAAGGTTGTGC 2864

Seq_name:    [illegible]

seq_documentation_block:
; Sequence 1, Application PC/TUS9307406
;  GENERAL INFORMATION:
;     APPLICANT:  MARGOLIS, Richard U.
;     APPLICANT:  RAUCH, Uwe
;     APPLICANT:  MARGOLIS, Renee K.
;     TITLE OF INVENTION:  CLONING, EXPRESSION AND USES FOR A
;     TITLE OF INVENTION:  NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
;     NUMBER OF SEQUENCES:  48
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Browdy and Neimark
;       STREET:  419 Seventh Street, N.W.
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  U.S.A.
;       ZIP:  20004
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/07306
;       FILING DATE:  03-AUG-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/922,911
;       FILING DATE:  04-AUG-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Townsend, Guy K.
;       REGISTRATION NUMBER:  34,033
;       REFERENCE/DOCKET NUMBER:  Margolis IA PCT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-628-5197
;       TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5191 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
;     FEATURE:
```

```
;       NAME/KEY:  CDS
;       LOCATION:  77..4847
PCT-US93-07306-1


alignment_scores:
            Quality:   95.00              Length:    401
              Ratio:   0.565                Gaps:     27
 Percent Similarity:   41.895   Percent Identity: 24.439

alignment_block:
 US-09-126-945B-2 x PCT-US93-07306-1   ..

 Align seq 1/1  to: PCT-US93-07306-1  from: 1  to: 5191

       5 SerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProProAs 21
         ||||||:::|||      :::|||||||||   :::         ||||:
    1766 TCCCCGTCGCTG......ATTTCACCCAGTACTGTC.........CCCAG 1800

      21 pThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGly. 37
         :|||   |||      |||||:::   :::|||      |||   :||
    1801 CACTGACAGTACTCCTGGCCTGAAGCCAGGGGCAGATGAGGCCCCTGGTG 1850

      38 ...........LeuGluArgArgAspTrpSerProSerProProAlaThr 53
                   :::::::::      |||   ||||||   |||||||:::
    1851 TGAAGTCAGCCATCCACCACCCACCTGGTTGCCCTCAGAACCCGCTGTC 1900

      54 ...........ProGluGlnGlyLeuSerAla................. 57
                   ||||:::::::::|||||||||
    1901 CCATCCTCCATTCCCTCAGAGGCCCTAAGTGCTGTCTCCCTACAGGCATC 1950

      57 ..................................................... 57

    1951 CCCTGGTGATGGCTCTCCAGACTTCCCCATTGTAGCCATGCTTCGAGCCC 2000

      58 .....PheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSer.... 71
           :::::|||        :::   |||  |||:::   |||
    2001 CCAAACTGTGGCTTCTGCCACACTCTACACTCGTCCCGAATGTGTCCCTA 2050

      72 .........SerTrpAlaAlaLysAlaProGlyAlaSerSerArgGlu.. 84
                 |||  |||:::     |||::::::     ::::||
    2051 ATCCCACTCTCCCCAGCCTCTCCACTCCCCTCCTCGGTCCCAGAAGAACA 2100

      85 ..............................GluProProGluGluProGlyG 92
                                     |||  |||:||  |||  :
    2101 GGCTGTCAGACCTGTCAGCTTTGGAGCAGAAGACCCCGAGACCCCATTTC 2150

      92 In.....................................CysProValIle 99
         ||                                    |||
    2151 AGACCACCATGGCTGCCCCAGGTGAAGCCAGCCACGGATCCCCTGAGGCA 2200

      97 AspSer................................................ 98
         ||||||
    2201 GACTCCATAGAAATCGAGGGGATCAGCTCCATGCAGGCTACAAAGCACCT 2250

      99 ....GlnAlaProAlaGlySerLeuAspLeu..................... 107
             :::::::|||   :::||||||||||
    2251 CATCTCTGGCCCATGGGCTTCTTTGGACTCCAGTAATGTGACAGTGAATC 2300

     108 ..ValPro.........GlyGlyLeuThrLeuGluGluHisSerLeuGlu 125
           ||||||         |||   |||      |||:::      |||:::
    2301 CTGTCCCTTCTGATGCTGGCATCCTAGGGACTGAGTCTGGGGTCTTGGAC 2350

     121 GlnValGlnSerMetValVal...GlyGluValLeuLysAspIleGluTh 136
                 |||   :::      |||:::::      |||:::
    2351 TTACCAGCGACTCCCACATCACAGCGCACAGCCCACTCTCTCACATCCTCT 2400

     136 rAlaCysLysLeu.................LeuAsn.IleThrAlaAsp 146
           |||     |||                 ||||:: :::::   :::
    2401 GGCCACCTGGCTACCACTGCCTGGCCACGGACTGGACACTGGCTCCCAGT 2450

     147 ProMetAspTrpSerProSer                    AsnValGlyy 157
         |||      |||:::|||                      ||||:::|||
```

```
    2451 CCACACCCATCAAACCCATCCACTAACCACCAGTCTCCAAGCCTACAC 2498

     157 sTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLys. 173
           |||   ||||||   |||            :::||||||      :::
    2499 .TGG...CTTGGAAGCAG...........GTGCCACCAAAGAGCCAAT 2542

     174 AlaPheGlnGluLeuAlaGlyLysGluLeuCys.............AlaMe 186
         :::   ::::::   :::|||   :::   |||             :::::
    2533 GGAGGCCACCATGGATGTGGTCCCCCAGCACTGTTGATGCCACTTCGGGGT 2582

     186 tSerGlnGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeu. 202
         :      ::::::|||:::          |||   |||   |||
    2583 CTGAACCCAAAAGTTCCATTTCTAGCACCCATGTGGTTGTGACTGCCAGCT 2632

     203 ...........HisAlaHisLeuAspIleTrpLys.SerAlaAlaTrpMe 215
                   |||::::|||   |||   :::||| |||   :::   ::
    2633 GGGGACCAGGGCACACCCACACTGACCCCTACAAGCTCTGAAGGTCAGGT 2682

     215 tLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerG 232
         :      :::   |||   |||::::::          ::::||      :
    2683 GGTGGCCCAGGAGTCACTGGGAACCCTC.........ACCAGTCTGCCTT 2723

     232 luGluSerTrpThrAsp..............SerGluValAspSerSer 243
         :::::   |||::::::                :::||||||   |||
    2724 CTCATCCCTGGTCATCTCTGGCCTCCAGCATGGACGAAGTGGCCTCGGTT 2773

     244 CysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeu.I 260
           |||||:::|||   :::|||||:::::::   :::   ||||: |
    2774 TCCTCAGGAGAACCCACAAGCTTGTGGGACATCCCCAGCACTCTGATACC 2823

     260 euLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLys 276
         ||   :||   :::                 ||||:|||
    2824 TGTGTCCTTGGGCT..................TGGATGAAT......... 2866

     277 GlyIlePheLysIleGluAspSerAlaGluValAlaArgLeuTrp 291
                                    |||:::   ||||||||
    2867 ..........................CAGACCTGAAGGTTGTGG 2864

seq_name: [illegible] Data [illegible] Mis [illegible] US-08-938,291A-2

seq_documentation_block:
; Sequence 2, Application US/08938291A
; Patent No. 6117674
;  GENERAL INFORMATION:
;    APPLICANT:  Lev, Sima
;    APPLICANT:  Plowman, Gregory D.
;    APPLICANT:  Schlessinger, Joseph
;    TITLE OF INVENTION:  RDGR PROTEINS AND RELATED
;    TITLE OF INVENTION:  PRODUCTS AND METHODS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/938,291A
;      FILING DATE:  September 26, 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,337
;      FILING DATE:  October 11, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
```

```
;       REGISTRATION NUMBER:  42,427
;       REFERENCE/DOCKET NUMBER:  228/172
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489 1600
;       TELEFAX:  (213) 955 0440
;       TELEX:  67 3510
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4190 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US 08 938 291A 2


alignment scores:
            Quality:   94.50              Length:    290
              Ratio:  0.768                Gaps:     17
 Percent Similarity:  42.414   Percent Identity: 25.172

alignment_block:
 US 09 126 945B 2 x US 08 938 291A 2   ..

 Align seq 1/1  to: US 08 938 291A 2  from: 1  to: 4190

        4 SerAlaSerProGlyLeuSerSerValSerProSer..............  14
          ||||||:::  |||  :::|||  :::  |||
     1905 AGTGCTAACGCGGGCACCGGAGTCGGGGCAGCAGCCGCCGTGGGAGCAT 1954

       15 .......HisLeuLeuLeuProProAspThrValSerArgThrGlyLeuG  29
                :::|||||  |||              |||     |||
     1955 GAACAATGAGCTGCTCTCTCCGGAGTTTGGCCCAGTGCGGGACCCCCTG. 2003

       29 IuLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerPro  45
             |||  |||:::  ||||||  |||          ||||||
     2004 .......GCAGATGGTCTCCAAGGCCTGGCTCCC.....CGCAGCCCA 2039

       46 SerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTy  62
          :::|||  |||  |||  |||  :::  :::
     2040 GAACCCTCGGCCTTGCCTCCCCAGCGCATCCCCAGC.............. 2075

       62 rPheAspMetLeuTyrProGlu.............................  69
            ||||||     ||||||
     2076 ....GACATGGCCAGTCCTGAGCCCGAGGGCTCTCAGAACAGCCTTCAGG 2125

       70 ..............AspSerSerTrpAlaAlaLys...............  76
                       |||||||||      :::
     2126 CAGCCCCGCAACCACCTCCTCCTGGGAGCCCCGGCGGGCAAGCACGGCC 2175

       77 ......AlaProGlyAlaSerSerArgGluGluProProGluGluProGl  91
                |||:::||||||||      |||  |||:::  |||::
     2176 TTCTGCCCACCCCCTGCCAGTTCC.....GAGGCACCTGACGGCCCCAG 2225

       91 uGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuV 109
          ::::  :::|||  ::::::  :::      |||
     2226 CAGCACTGCCCGGCCTTGACTTCAAGGTCTCTGCCTTCTTCCTCTTCGGCT 2275

      109 alProGlyGlyLeuThrLeuGluGluHis..............SerLeu 119
            |||  ||||||:::|||      :::              :::|||
     2276 CCCCACTCGGGCCTGCTGCTGGCTCTGCCCAAAACTGTGAATGCCCGCCTG 2325

      120 GlnGlnValGlnSerMetValValGlyGluValLeuLysAspIleGluTh 136
          |||  :::|||                              ::::::
     2326 GAGGCAGCCCAG..............................ATGCGCCC 2365

      136 rAlaCys...LysLeuLeuAsnIle......ThrAlaAspProMetAspT 150
          ||||||  ::::::  |||:::      :::||||||||||
     2366 AGCCTGTGAACAGATCTACAACCTCTTCCACGCGGCCGACCCCTGC.... 2381

      150 rpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyr 166

     2382 ..................................................GCCTCA 2387
```

```
      167 ArgLeuProPro....MetGlyLysAlaPheGlnGluLeuAlaGlyLysGl 182
          ||||||    |||   ::::::      ||||||   :::|||
     2388 CGCCTCGAGCCCCTGCTGGCCCCGAAGTTCCAGGCCATCGCC........ 2429

      182 nLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGly. 198
                  ::::::      ::::::   ||::::   |||||||||
     2430 .......CCACTGACCGTGGCCCCGTAC...CAGAAGTTCCCCCTGGGAG 2469

      199 .....................GlyAspValLeuHisAlaHisLeuAspIle 208
                               :::|||:::|||::::::|||   ::::::
     2470 ATGGCTCATCCTGATGTGGCCGACACTCTGCAGACGCACTCTAGCCTC 2519

      209 TrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIleHi 225
          :::   :::          |||          ::::::|||
     2520 TTCTGGAGGAGCTGGAGATGCTGGTGCCCTCAACACCC............ 2558

      225 sTyrCysAlaSerThrSerGluGlnSerTrpThrAspSerGluValAspS 242
              :::|||||||||          |||     ||||||:::  :
     2559 .......ACCTCTACTAGCGGTGCTTCTGGAAGGGCAGTGAGTTGGCCA 2601

      242 erSerCysSerGlyGlnPro 248
          :::::      :::||||||
     2602 CTGACCCCCCGGCCCAGCCA 2621

<rej_name  /cre1_5/ptodata/1/iia/US_CoMB.seq:US 08 785 810A 1

seq_documentation_block:
; Sequence 1, Application US/08785810A
; Patent No. 5840532
;  GENERAL INFORMATION:
;    APPLICANT:  McKnight, Steven L.
;    APPLICANT:  Russell, David W.
;    TITLE OF INVENTION:  Neuronal PAS Domain Protein
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/785,810A
;      FILING DATE:  21-JAN 1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UTSD:1226
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2078 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US 08 785 810A 1


alignment_scores:
            Quality:   94.00              Length:    338
              Ratio:  0.648                Gaps:     19
 Percent Similarity:  42.899   Percent Identity: 24.852

alignment_block:
```

```
US-09-126-945B-2 x US-08-785-410A-1   ..

Align seq 1/1  to: US-08-785-410A-1  from: 1  to: 2078

      12 SerProSerHisLeuLeuLeuProProAspThrValSerArgThrGlyIe  28
         :::|||:::::   |||   ||||||      |||   |||:::|||
     444 GCGCCICAGCGTCACCTACCTCCGCCTGCGCCGGTTCGCCGCGCTGGT.. 491

      28 nGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrp...S  44
            ||||||   ||||||   :::          :::|||
     492 ....GGCGCCGCCCTGGGGGCTGAGAGCCGCGGGGCCGCCAGCTGGCCTC  537

      44 erProSerProProAlaThrProGluGlnGlyLeu.SerAlaPhe.....  58
         |||   ||||||         :::   ||||||:::::::::
     538 GCGCCAGGCCGCCGCGGCCCGCAGCGCTGGTCTCCGAAGTCTTCGAGCA  587

      58 ..................................................  58

     588 GCAGCTGGGAGGTCACATCTTGCAGTCCTGGATGGCTTTGTGTTCAGCT  637

      59 .....................................................Tyr  59
                                                             |||
     638 TGAACCAGGAAGGAAAATTCCTCTACATCTCAGAGACAGTCTCCATCTAT  687

      60 Leu.............................SerTyrPheAspMetIe  66
         |||                           |||   ||||||   ::
     688 CTGGGTCTCTCACAGGTGGAGATGACGGGCAGCAGCGTCTTCGACTACAT  737

      66 nTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerA  85
         :::::||   |||   |||   :::   :::      |||         |
     738 TCACCCTGGGGACCACTCAGAGGTGCTGGAGCAACTGGGCGCTG.....C  781

      86 rgGluGluProProGluGluProGluGlnCysProValIleAspSerGln  99
         ||      |||      |||      |||   :::|||:::
     782 GGACGACGACGCCCGGCCCCCCAACCCCGTCCTCCGTCTCCTCTTCCTCC  831

     100 AlaProAlaGlySerLeuAspLeuValPro.........GlyGlyLeuGlu  113
         :::   :::::|||||      :::|||          ::::|||||
     832 TCCTCTTCCTCTTCGCTTGCAGATACCCCCGAGATCGAGGCCAGCCTCAC  881

     114 r...LeuGluGluHisSerLeuGluGlnValGlnSerMetValVal..Gl  128
         |   :::         ||||||   |||   :::|||:::   |||
     882 CAAGCTGCCCCCCTCCTCCCTGGTCCAGGAGCCGCTCCTICTTGTCCCCA  931

     128 yGluValLeuLysAspIleGluThrAlaCysLysLeu...LeuAsnIle1  144
          |||:::          |||   |||:::::::      |||:::::::
     932 TGAAATCCACGCTCACCAAGAGGGGGCTGCACGTCAAGGCCTCAGGGTAC  981

     144 hrAlaAspProMetAspTrpSerProSerAsnValGlnLys.IrpLeuIe  160
         :::   |||   |||||||:::|||||:::   :::        ||||:
     982 AAGGTCATCCACGTGACTGGGCGCCTTCGGGCCCACGCCTGGGCCTTGT  1031

     160 nTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnC  177
         :        |||         |||||||||
    1032 GGCCCTCGGGCAC......ACGTTGCCCCCG.................... 1056

     177 luLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGluPheArgGln  193

    1056 ..................................................  1056

     194 ArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLy  210
            ::: |||||::         |||||||:::|||:::   ::::::::
    1057 ...GCCCCCCTGGCTGAGCTGCCACTCCATGGACACATGATCGTCTTCCG  1106

     210 s.SerAlaAlaIrpMetLysGluArgThrSerProGlyAlaIleHisTyr  226
         : |||   |||||:::            |||||||::   :::
    1104 TCTCA...GCCTGGGTC............TCACCATCCTTGCTTGTGAGA  1138

     227 CysAlaSerThrSerGluGluSerIrpThrAspSerGlu..ValAspSer  242
         :::   ||:::::::      ::|||      ::::::  |||
    1139 GCAGAGTCAGCGACCACATGGACCTGGGGCCCTCAGAGCTGGTGGGCCGC  1188

     243 SerCysSerGlyGlnProIleHis.....................IenIrpGl  258

         ||||||      |||   :::|||                    :::   ||
    1189 AGCTGCTAC...CAGTTTGTCCACGGACAAGACGCCACGAGGATCCGCCA  1235

     258 nPheLeuLysGluLeuLeuLysProHisSerTyrGlyArgPheIleA  270
         |          ::|||||||   |||   :::             :::   |
    1236 GAGCCACGTGGACTTGCTGGACAAGGGTCAGGTGATGACTGGTTACTACC  1285

     270 rgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGln  286
         ||||||||:::::      |||:::      :::::::::   |||
    1286 GTTGGCTGCAGCGTGCCGGGGGCTTCGTGTGGCTGCAGTCTGTGGCCACA  1335

     287 ValAla 288
         ||||||
    1336 GTGGCT 1341

Seq_name: /tmp2_f/pro1r1_1/Seq_5A_COMB.seq:US-08-344-428-1

seq_documentation_block:
; Sequence 1, Application US/08344428
; Patent No. 5665586
;  GENERAL INFORMATION:
;    APPLICANT:  Nakamura, Etsuo
;    APPLICANT:  Isuzuki, Hiroshige
;    APPLICANT:  KitadoKoro, Kenjo
;    APPLICANT:  Shin, Nasaru
;    APPLICANT:  Teraoka, Hiroshi
;    TITLE OF INVENTION:  No. 5665586el Protease
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  James F. Haley, Jr., Fish & Neave
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10020-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 720Kb storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS Dos 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/344,428
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/00592
;      FILING DATE:  30-APR-1993
;      APPLICATION NUMBER:  JAPAN 4 126511
;      FILING DATE:  19-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley, Jr., James F.
;      REGISTRATION NUMBER:  27794
;      REFERENCE/DOCKET NUMBER:  SHON-7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)596-9000
;      TELEFAX:  (212)596-9090
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2064
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces fradiae
;    FEATURE:
;      NAME/KEY:  -35 signal
;      LOCATION:  459..464
;      IDENTIFICATION METHOD:  by experiment
;      NAME/KEY:  -10 signal
;      LOCATION:  482..488
;      IDENTIFICATION METHOD:  by experiment
;      NAME/KEY:  CDS
```

```
;       LOCATION:  435..1505
;       IDENTIFICATION METHOD:  by experiment
;       NAME/KEY:  sig_peptide
;       LOCATION:  435..944
;       IDENTIFICATION METHOD:  by experiment
US-08-343-428-1


Alignment scores:
            Quality:   93.50              Length:     435
              Ratio:   0.786               Gaps:      18
 Percent Similarity:  45.522   Percent Identity: 20.896

Alignment block:
 US-09-126-945B-2 x US-08-343-428-1  ..

 Align seq 1/1  to: US-08-343-428-1  from: 1  to: 2064

        4 AlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProPr  20
          |||:::|||:::     |||:::||||||||||                ||
     1276 GCGGCACCGTCAGCGCCGTCAACGTCACCGTCAACTACAGCGACGGCCCC 1325

       20 oAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValG  37
          |:::|||:::
     1326 GICTACGGCA........................................ 1335

       37 lyLeuGluArgArgAspTrpSerProSerProProAlaThrProGluGln  53
                    ||||||      ||||||||||   |||
     1336 ..............TGGTCCGCACGACCGCCTGCTCCGCCGGCGGCG 1368

       54 GlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAs  70
            ::|||                      :::      |||:::
     1369 ACAGCGGCGGCG.......CCCACTTCGCCGGTTCCGTCGCCCTCGGC 1409

       70 pSerSerTrpAlaAla.................lysAlaProGlyAlaS  81
          |||:::   ||||||                 ||||||   ::::
     1410 ATCCACTCGGGCAGCTCCGGCTGCACGGGCACCAACGGCTCCGCCATCCA 1459

       81 erSerArgGluGluProProGluGluProGluGlnCysProValIleAsp  97
          :::||||||:::      |||      |||                 |||
     1460 CCAGCCGGTCCGCGAGGCCCTCAGCGCCTATGCCGTGAACGTGTACTGAC 1509

       98 SerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLe 114
                |||||||||
     1510 CCGCCCCCGCCCGGCCGGG.................................. 1537

      114 nGlnGluHisSerLeuGluGlnValGlnSerMetValValGlyGluValL 131
                                  :::|||   :::||||||   :::
     1538 .......................CGGAGTGCCACAGTGGGCGGTATGA 1562

      131 euLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspPro 147
                 ::::::|||||||
     1563 GTGATCGTCTCCATACCGCC.................................. 1592

      148 MetAspTrpSerProSerAsnValGluLysTrpLeuLeuTrpThrGluHi 164
                                                          ||
     1593 ..................................................CA 1594

      164 sGlnTyrArgLeuProPro.....................................MetG 172
          |   |||::|||   |||                                    :::
     1595 CACCTACGAGCTGCCACCGGCCGTGCTCGCCTCGATCCGCGCCCTCCTCC 1644

      172 lyLysAlaPheGlnGluLeuAlaGlyLysGlnLeuCysAlaMetSerGlu 188
               ||||||:::                                 :::|||||
     1645 AGCACGCCTTCGAAGGC.....................GACCTGACGCAG 1693

      189 GluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHi 205
          |||:::::::::::::          |||||||||       :::|||||::
     1694 GAGGACTGGGACGCAC......GCCCTCGGCGGC......GTCCACGCGTA 1691

      205 sLeuAsp.....................IleTrpLysSer.............. 211
          :   :::                  ||| |||:::::
     1692 CGCCGACGACCCCGGCGGCGCGGATCGTGGCGCACGGCAGCGTCGTCCATGC 1741

      212 AlaAlaTrpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAl 228
          |||||||||           ::::::||||||                ||
     1742 GCGGCGTGG.............TCCACGCCGGCC.............GC 1764

      228 aSerThrSerGluGluSerTrpThrAspSerGluValAspSerSerCysS 245
          |   |||::::::::   :::|||
     1765 GCCCACCGCGTCGGGTACGTGG.............................. 1786

      245 erGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLys 261
                                                          :::
     1787 ..................................................AGG 1789

      262 ProHisSerTyrGlyArg.................................. 267
          |||   :::||||||||
     1790 CCGTGGGCGTACGGGCGGACCGCCGCCGCCAGGGGCTGGGCACGCGGGTG 1839

      268 .PheIleArgTrpLeuAsnLys.GluLysGlyIlePheLysIleGluAsp 283
           :::   ||||||   :::::: ::::::|||:::
     1840 ATGGCGGCGCTGGAGGAGGTCCTCGACGGGGCGTA.................. 1874

      284 SerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMe 300
                ::::::   |||         :::|||::::::||||||||||
     1875 ......CGACTCGCCGCGCTCTCCGCCTCCGACGCGGCGCCTGCTC 1918

      300 t 300
          :
     1919 A 1919

Seq_name: /cgi-data/pl-data/tmp_SAMPLE_seq_US-08-435-609-4

seq_documentation block:
; Sequence 4, Application US/08435609
; Patent No. 5955460
;  GENERAL INFORMATION:
;     APPLICANT:  Chamberlain, Jeffrey S.
;     APPLICANT:  Amalfitano, Andrea
;     APPLICANT:  Hauser, Michael A.
;     APPLICANT:  Kumar Singh, Rajendra
;     APPLICANT:  Bartlett O'Connor, Dennis J.
;     TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Medlen & Carroll, LLP
;       STREET:  220 Montgomery Street, Suite 2200
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  United States Of America
;       ZIP:  94104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/435,609
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Ingolia, Diane E.
;       REGISTRATION NUMBER:  40,027
;       REFERENCE/DOCKET NUMBER:  UM-02484
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  44303 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "DNA"
```

```
US-08-735-609-4

alignment_scores:
            Quality:   93.50              Length:    366
              Ratio:  0.592               Gaps:       18
 Percent Similarity:  43.169   Percent Identity: 23.497

alignment_block:
 US-09-126-945B-2 x US-08-735-609-4    ..

 Align seq 1/1  to: US-08-735-609-4  from: 1  to: 34303

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa  46
         :||         |||||||||||||   :::   ||||||   |||
   12429 CCCCGGTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 12470

      36 lGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu. 52
          ||||||      |||         |||                 ||||||
   12471 .GGTTGCCTCCCCGTCAPGCAAGACCCCGCTTGCAAATTCCTCCGCAAA 12519

      53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl  69
         ||||||   :::   |||                 ::||||:|||||:
   12520 CAGGGACGAGCCCCTTTT.................TTGCTTTCCCAGA 12561

      69 uAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGlu  85
         :  |||:::  ||||||  ||||||  :::|||:::|||:|||
   12562 TGCATCCGGTGCTGCGGCAGATGCGCCCCCCTCCTCAGCAGCGGCAAGAG 12611

      86 ProProGluGluProGluGlnCysProValIleAspSerGluAlaProAl 102
            :::|||  :::|||          |||     |||::
   12602       CAACAGCACCGGCACACATCCAGCGGCACCCCTCCCTCCTCCTAC 12645

     102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
         :::::||              ||||||   |||
   12646 CGGGTCA.............GGAGGGGCGACA................. 12684

     119 euGluGlnValGlnSerMetValValGlyGluVal............... 130
            :::  |||:::::::  :::  ||||::
   12685 ..TCCGCCTTCACCCCCAGCACATGGTCATTACGAACCCCCGCGCCC 12732

     131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsn11 143
                     |||:::|||  :::   |||  :::::
   12713 CGGGCCCGGCACTACCTGCACTTCGAGCAGCGGCAGCGCCTGCTCGCGCGT 12762

     143 eThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuL 160
         :  |||             ||||||:::::      :::      |
   12763 AGGAGGCCC............TCTCCTGAGCCGCTACCCAAGGGTGCAGC 12809

     160 euTrpTrpGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGln 176
         ||      :::   :::             |||      :::   ||:::
   12801 TGAAGCGTCATACCGCTGAGCCGTACGTCCCGCCGCAGAACCTGTTCCGC 12850

     177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGl 193
         :::     |||:::||          |||   ::::::::::||:::
   12851 GACCGCGAGCGAGACGACCCCGACGACGATGCGGATCGAAAGTTCCACCGC 12900

     193 n............ArgSerProLeuGlyGlyAspValLeuHisAlaHis 209
                     |||      ||:::   :::   |||      :::
   12901 AGGGTGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGG 12950

     209 euAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGly 225
         |||:::      :::|||          :::          ||||||:::
   12951 AGCACTTTGAGCCCGACGCG.........CGAACCGGGATTAGTCCCGCG 12994

     225 AlaIleHisTyrCysAlaSer.........ThrSerGluGluSerTrpTh 236
            |||   :::|||:::          |||:::   |||:::   :::
   12995 CGCGCACACGTGGCGGCGCCGACCTGGTAACCGCATACGAGCAGACGGT 13041

     236 rAspSerGluValAspSerSerCysSerGlyGlnPro.............. 248
         ::::::::|||:::::   :::   |||   :::
   13042 CAACCAGGACATTAACTTTCAAAAAAGCTTTAACAACGCACGCGCTACGC 13091

     249 ..............................IleHisLeuTrpGlnPhe 264
                                        :::||||||||||:::|||
   13092 TTGTGGCGCGCGAGGAGGTGGCTATAGGACTGATGCATCTGTGGGACTTT 13141

     255 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle..... 269
         ::::::   |||       :::|||:::|||          :::
   13142 GTAAGCGCGCTGGAGCAAAACCCAAATAGCAAGCCGCTCATGGCGCAGCT 13191

     269 ..................................................... 269

   13192 GTTCCTATAGTGCAGCACAGCAGGACAACGAGGCATTCAGGGATGCGC 13241

     270 ............................ArgTrpLeu.............Asn 278
                                      ||||||||| |||
   13242 TGCTAAACATAGTAGAGCCCGAGGCCCGCTGGCTGCTCGATTTGATAAAC 13291

     271 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgIle 290
         ::::::|||       |||   |||   :::::|||
   13292 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 13336

     290 uTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerA 307
                    :::    ||:::::|||||:::   |||
   13337 .............GACAAGGTGGCCGCCATCAACTATTCCATGCTT.... 13379

     307 rgSerIleArgGluTyrTyrLysLysGlyIleIleArgLysPro 329
           |||:::   :::::|||   :::   |||   :::   |||
   13370 ..AGCCTGGGCAAGTTTTACGCCCGCAACATATCCATACCCCT 13411

seq_name: /cgn2_6/ptodata/1/iia/*B_COMP.seq:US-08-735-609-4

seq_documentation_block:
; Sequence 4, Application US/08735609
; Patent No. 5994132
;   GENERAL INFORMATION:
;        APPLICANT: Chamberlain, Jeffrey S.
;                   Amalfitano, Andrea
;                   Hauser, Michael A.
;                   Kumar-Singh, Rajendra
;                   Bartlgan-O'Connor, Dennis J.
;        TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
;        NUMBER OF SEQUENCES: 15
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll, LLP
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC DOS/MS DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/735,609
;             FILING DATE: 23-Oct-1996
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ingolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: UM-02484
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 34303 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;             DESCRIPTION: /desc = "DNA"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
```

```
US 08 735 609 4


Alignment_scores:
            Quality:   93.50              Length:     366
              Ratio:   0.592                Gaps:      18
 Percent Similarity:  43.169   Percent Identity: 23.497

Alignment block:
 US 09 126 945B 2 x US 08 735 609 4   ..

 Align seq 1/1   to: US 08 735 609 4   from: 1   to: 44303

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa  36
         |||          ||||||||||||   :::   ||||||   |||
   12429 CCCCGGTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 12470

      36 IGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu  52
         ||||||      |||          |||                ||||||
   12471 .GGTTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAA 12519

      53 GluGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl  69
         ||||||   :::   |||                 :::|||:::|||::
   12520 CAGGGACGAGCCCCTTT.................TTGCTTTCCCAGA 12561

      69 nAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGlu  85
         :   ||||::   ||||||   ||||||   :::|||::: |||:::|||
   12562 TGCATCCGTTGCTGCGGCAGATGCGCCCCCCTCCTCAGCAGCGGCAAGAG 12611

      86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
         :::|||   :::|||                ||| |      ||||:
   12612 .....CAAGAGCAGCGGCAGACATGCAGGGCACCCTCCCCTCCTCCTAC 12665

     102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
         :::::|||              ||||||   |||
   12666 CGCGTCA.............GGAGGGGCGACA................. 12704

     119 euGluGlnValGlnSerMetValValGlyGluVal............... 130
            :::   |||::::::   :::   |||:::
   12705 ..TCCGCGCTTCGACGCGGCAGCAGATGGTGATTACGAACCCCCGCGGCGC 12754

     131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIl 148
                         |||:::|||   :::      |||   :::::
   12755 CGGGCCCGCACTACCTGGACTTGGAGGACGGCCGAGGGCCTGGCGCGGCT 12762

     149 eThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrpLeuL 166
         :   |||                 ||||||:::::      :::      |
   12763 AGGAGCGGCC............TCTCCTGACGCGGTACCCAAGGGTGCAGC 12808

     167 euTrpThrGluHisGluTyrArgLeuProProMetGlyLysAlaPheGlu 186
         ||      :::   :::                |||    :::    |||:::
   12809 TGAACGCGATATCGCGTGAGGCGTACGTGCCGCGCAGAACCTGTTTCGC 12858

     187 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGluPheArgSe 206
         :::      |||:::|||          |||   ::::::::::|||:::
   12859 GACCGCGACGGAGAGGAGCCCGAGGACATGCGGGATCGAAAGTTCCACGC 12908

     207 n.............ArgSerProLeuGlyGlyAspValLeuHisAlaHisL 226
                       |||      |||:::   :::   |||      :::
   12909 AGGGCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGG 12958

     227 euAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGly 242
         |||:::      :::|||          :::          ||||||::
   12959 AGGACTTTGAGCCCGACGCG.........CGAACCGGGATTAGTCCCGCG 12999

     243 AlaIleHisTyrCysAlaSer.........PhrSerGluGluSerIrpHi 260
            |||   :::|||:::            |||:::   |||:::   ::
   13000 CGCGCACACGTGGCGGCCGCCGACCTGGTAACCGCATACGAGCAGACGCT 13049

     261 rAspSerGluValAspSerSerCysSerGlyGluPro.............. 278
         :::::::|||:::::::   :::   |||   :::
   13050 GAACCAGGAGATTAACTTTCAAAAAGCTTTAACAACCACGTGCGTACGC 13099

     279 .....................IleHisLeuIrpGluPhe 294
                              :::||||||||||::|||
   13100 TTGTGACGCGCGAGGAGGTGCTATAGGACTGATGCATCTGTGGGACTTT 13149

     295 LeuLysGluLeuLeuLysProHisSerTyrGlyArgPheIle..... 309
         ::::::   |||     :::|||:::|||      :::
   13150 GTAAGCGCGCTGGAGCAAAACCCAAATAGCAAGCCGCTCATGCGCAGCT 13199

     309 ..................................................... 309

   13200 GTTCCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 13249

     310 .........................ArgTrpLeu.............Asn 323
                                 ||||||||              |||
   13250 TGCTAAAcATAGtAcAGcCCGAGGGCGcTGCTGCTGATTTGATAAAC 13299

     324 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLe 340
         ::::::|||          |||   |||   :::::|||
   13300 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 13346

     340 nTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuIleA 357
                        :::      |||:::|||||||:::   |||
   13347 .............GACAAGGTGGCCGCCATCAACTATTCCAAGCTT.... 13389

     358 rgSerIleArgGluTyrTyrLysLysGlyIleIleArgLysPro 421
         |||:::   :::::|||   :::   |||   :::   |||
   13390 ..AGCCTGGGCAAGTTTACGCCCGCAAGATATACCATACCCCT 13431

seq_name: /cgn2_6/ptodata/2/ina/6A_COMP/seq_US_09_415_472_4

seq_documentation block:
; Sequence 4, Application US/09415472
; Patent No. 6057158
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Bauser, Michael A.
;    APPLICANT:  Kumar Singh, Rajendra
;    APPLICANT:  Hartigan O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States Of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/415,472
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/735,609
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ingolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM 02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705 8410
;      TELEFAX:  (415) 397 8338
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44303 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

```
;     MOLECULE TYPE:  other nucleic acid
;     DESCRIPTION:  /desc   "DNA"
US-09-415-372-4


alignment_scores:
            Quality:   93.50              Length:     366
              Ratio:   0.592                Gaps:      18
 Percent Similarity:   43.169   Percent Identity: 23.497

alignment_block:
 US-09-126-945B-2 x US-09-415-372-4   ..

 Align Seq 1/1  to: US-09-415-372-4  from: 1  to: 34303

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 36
         |||          |||||||||||||   :::   ||||||   |||
   12429 CCCCGGTTCGACTCTCGGACCGGC...GGGACTGCGGCGAACGGG..... 12470

      36 lGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu. 52
          ||||||      |||         |||              ||||||
   12471 .GGTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGCAAA 12519

      53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl 69
         ||||||   :::   |||                  :::|||:::|||:
   12520 CAGGGACGAGCCCCTTTT.................TTGCTTTTCCCAGA 12551

      69 nAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGln 85
         :   ||||::   ||||||   ||||||   :::|||::: |||:::|||
   12552 TGCATCCGGTGCTGCGGCAGATGCGCCCCCCTCCTCAGCAGCGGCAAGAG 12601

      86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
            :::|||   :::|||                |||      |||::
   12602 ......CAAGAGCAGCGGCAGACATGCAGGGCAGCCTCCCCTGCTGCTAC 12645

     102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
         :::|||                |||||   |||
   12646 CGCGTCA..............GGAGGGGCGACA................. 12664

     119 euGluGlnValGlnSerMetValValGlyGluVal............... 130
           :::   ||||::::::   :::   ||||::
   12665 ..TCCCCGTTGACGCGGCAGCAGATGGTGATTACGAACCCCCGCGGCGC 12712

     131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsuI 143
                     |||:::|||   :::      |||   :::::
   12713 CGGGCCCGGCACTACCTGGACTTGGAGGAGGGCGAGGGCCTGGCGCGGCT 12762

     143 eThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuL 160
         :   |||                |||||:::::::      :::      |
   12763 AGGAGCGCCC............TCTCCTGAGCGGTACCCAAGGGTGCAGC 12809

     160 euTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheIle 176
         ||    :::   :::                |||      :::   ||:::
   12801 TGAAGCGTGATACGCGTGAGGCGTACGTGCCGCGGCAGAACCTGTTTGGC 12850

     177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGl 193
         :::      |||:::|||          |||   :::::::::|||:::
   12851 GACCGCCAGGGAGAGGAGCCCGAGGAGATGCGGGATCGAAAGTTCCACGC 12900

     193 n............ArgSerProLeuGlyGlyAspValLeuHisAlaHisL 206
                   |||      |||:::   :::   |||      :::
   12901 AGGGCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGG 12950

     206 euAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGly 222
         |||:::    :::|||          :::              |||||||::
   12951 AGGACTTTGAGCCCGACGCG.........CGAACCGGGATTAGTCCGGGG 12997

     223 AlaIleHisTyrCysAlaSer.........ThrSerGluGluSerTrpIh 236
              ||   :::|||:::          |||:::   |||:::   ::
   12998 CGCGCACACGTGGCGGCCGCCGACCTGGTAACCGCATACGAGCAGACGGT 13047

     236 rAspSerGluValAspSerSerCysSerGlyGlnPro.............. 249
         :::::::|||:::::   :::   |||   :::

   13048 GAACCAGGACATTAACTTTCAAAAAAGCTTTAACAACCACGTGCGTACGC 13097

     249 ..............................IleHisLeuTrpGlnPhe 254
                                      :::|||||||||:::|||
   13098 TTGTGGCGCGCGAGGAGGTGGCTATAGGACTGATGCATCTGTGGGACTTT 13147

     255 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle..... 269
         ::::::   |||      :::|:|:::|||          :::
   13148 GTAAGCGCGCTGGAGCAAAACCCAAATAGCAAGCCGCTCATGCGCGCACT 13197

     269 ..................................................... 269

   13198 GTTCCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 13247

     270 ..........................ArgTrpLeu.............Asn 278
                                    |||||||||             |||
   13248 TGCTAAACATAGTAGAGCCCGAGGGCGCTGGCTGCTCGATTGATAAACC 13297

     274 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgIe 290
         ::::::::|          |||   |||   ::::::::|
   13298 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 13346

     290 uTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuIleA 307
                 :::       ||||::|||||::::   |||
   13347 .............GACAAGGTGGCCGCCATCAACTATTCCATGCTT.... 13389

     307 rgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
           ||:::   ::::::::||   :::   |||   :::   |||
   13390 ..AGCCTGGGCAAGTTTTACGCCCGCAAGATATACCATACCCCT 13431

seq_name= [redacted] MP seq_id=09-415-372 4

seq_documentation_block:
; Sequence 4, Application US/09244752
; Patent No. 6063622
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Hauser, Michael A.
;    APPLICANT:  Kumar-Singh, Rajendra
;    APPLICANT:  Hartigan-O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,752
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/735,609
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Impolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM-02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  34303 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc   "DNA"
US-09-244-752-4


alignment scores:
            Quality:   94.50              Length:     366
              Ratio:  0.592                Gaps:      18
 Percent Similarity:  43.169   Percent Identity: 23.497

alignment block:
 US-09-126-945B-2 x US-09-244-752-4   ..

 Align seq 1/1  to: US-09-244-752-4  from: 1  to: 84304

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa  36
         |||       ||||||||||||   :::   ||||||   |||
   12429 CCCCGGTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 12470

      46 lGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu.  52
         ||||||       |||          |||                |||||
   12471 .GGTTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAA 12519

      53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl  69
         ||||||   :::   |||                  :::|||:::|||:
   12520 CAGGGACGAGCCCCTTT.................TTGCTTTTCCCAGA 12551

      69 aAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGlu  85
         :  |||:::  ||||||  ||||||  :::|||::: |||:::|||
   12552 TGCATCCGGTGCTGCGGCAGATGCGCCCCCTCCTCAGCAGCGGCAAGAG 12601

      86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
            :::|||  :::|||           |||      |||::
   12602 .....CAACACCACGGGCAGACATCCACGGCACCCTCCGCTCCTCCTAC 12645

     102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
         :::|||              ||||||  |||
   12646 CGCGTCA.............GGAGGGCGACA................. 12664

     119 euGluGlnValGlnSerMetValValGlyGluVal............... 130
           :::  |||::::::  :::  |||:::
   12665 ..TCCGCGCTCACGCGGCAGCAGATGGTGATTACGAACCCCCGCGGCGC 12712

     131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIl 149
                     |||:::|||  :::    |||  :::::
   12713 CGGGCCCGGCACTACCTGGACTTGGAGGAGGGCGAGGGCCTGGCGCGGCT 12762

     143 eThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuL 160
         :  |||              ||||||:::::      :::       |
   12763 AGGAGCGGCC............TCTCCTGAGCGGTACCCAAGGGTGCAGC 12809

     160 euTrpThrGluHisGluTyrArgLeuProProMetGlyLysAlaPheGlu 176
         ||     :::  :::                |||     :::   ||:::
   12801 TGAAGCGTGATACGCGTGAGGCGTACGTGCCGCGGCAGAACCTGTTTCGC 12850

     177 GluLeuAlaGlyLysGluLeuLysAlaMetSerGluGluGlnPheArgSl 193
         :::      |||:::|||         |||  ::::::::::|||:::
   12851 GACCGCGAGGGAGAGGAGCCCGAGGAGATGCCGGATCGAAAGTTCCACGC 12900

     193 n............ArgSerProLeuGlyGlyAspValLeuHisAlaHisL 206
                     |||      |||:::  :::  |||      :::
   12901 AGGGCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGC 12950

     206 euAspIleTrpLysSerAlaAlaTrpMetLysGluArgIleSerProGly 222
         |||:::      :::|||          :::          |||||||:::
   12951 AGGACTTTGAGCCCGACGCG.........CGAACCGGGATTAGTCCCGCG 12991

     223 AlaIleHisTyrCysAlaSer.........ThrSerGluGluSerTrpIl 236
            |||  :::|||:::            |||:::  |||:::  ::
   12992 CGCGCACACGTGGCGGCCGCGACCTGGTAACCGCATACGAGCAGACGGC 13041

     236 rAspSerGluValAspSerSerCysSerGlyGluPro.............. 248
         :::::::|||:::::  :::  |||  :::
   13042 GAACCAGGAGATTAACTTCAAAAAAGCTTTAACAACCACGTGCGTACGC 13091

     249 ..................................IleHisLeuTrpGlnPhe 254
                                          :::||||||||||:::|||
   13092 TTGTGGCGCGCGAGGAGCTGGCTATAGGACTGATGCATCTGTGGGACTTT 13141

     255 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle..... 269
         ::::::   |||      :::|||:::|||        :::
   13142 GTAAGCGCGCTGGAGCAAAACCCAAATAGAAAGCCGCTCATGGCGCAGCT 13191

     269 ..................................................... 269

   13192 GTTCCTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 13241
         .
     270 ..........................ArgTrpLeu.............Asn 278
                                    |||||||||           |||
   13242 TGCTAAACATAGTAGAGCCCGAGGCCGCTGGCTGCTCGATTTGATAAAC 13291

     279 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLe 295
         ::::::|||       |||   |||   ::::::|||
   13292 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 13336

     296 uTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuIleA 312
                        :::       |||:::|||||||:::   |||
   13337 ..............GACAAGGTGGCCGCCATCAACTATTCCATGCTT..... 13369

     313 rgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
           |||:::   :::::|||   :::   |||   :::   |||
   13370 ..AGCCTGCGCAAGTTTTACGCCAAGATATACCATACCCCT 13411

<< parameters >>

seq_documentation_block:
; Sequence 4, Application US/09245497
; Patent No. 6084750
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Hauser, Michael A.
;    APPLICANT:  Kumar Singh, Rajendra
;    APPLICANT:  Bartlett-O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/245,497
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/735,609
;      APPLICATION NUMBER:  PCT/US97/01
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ingolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM-02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
```

```
;       LENGTH:  34303 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "DNA"
US-09-245-497-4


alignment_scores:
            Quality:   93.50              Length:    366
              Ratio:   0.592               Gaps:     18
 Percent Similarity:  43.169   Percent Identity: 28.497

alignment_block:
 US-09-126-945B-2 x US-09-245-497-4  ..

 Align_seq 1/1  to: US-09-245-497-4  from: 1  to: 34303

       20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 69
          |||          |||||||.|||||   :::   ||||||   |||
   12429 CCCCGGTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 12476

       86 lGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu. 5_
          ||||||      |||       |||                  ||||||
   12471 .GGTTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAA 12519

       53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl 99
          ||||||   :::   |||                 ::::|||||||||
   12520 CAGGGACGAGCCCCTTTT..................TTGCTTTCCCAGA 12561

       69 uAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGlu 85
          :   ||||::   ||||||   ||||||   :::|||:: |||||||:
   12552 TGCATCCGGTGCTGCGGCAGATGCGCCCCCTCCTCAGCAGCGGCAAGAG 12601

       86 ProProGluGluProGluGluCysProValIleAspSerGluAlaProAl 132
             ::|||   ::|||                |||      ||||:
   12602 .....CAAGAGCAGCGGCAGACATGCAGGGCACCCTCCCCTCCTCCTAC 12645

      102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
          :::|||                |||||   |||
   12646 CGCGTCA..............GGAGGGGCGACA.................. 12683

      119 euGluGluValGlnSerMetValValGlyGluVal............... 130
            :::   ||||||||:   :::   ||||::
   12665 ..TCCGCGGTTGACGCGGCAGCAGATGGTGATTACGAACCCCCGCGGCGC 12712

      131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIl 143
                      |||:::|||   :::      |||   :::::
   12713 CGGGCCCAGCACTACCTGGACTTGGAGGAGGGCGAGGCCTGGCGCGGCCT 12762

      144 eThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrpLeuL 160
          :   |||                |||||:::::::      :::     |
   12763 AGGAGCGCCC............TCTCCTGAGCGGTACCCAAGGGTGCAGC 12800

      161 euTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlu 176
          ||      :::   :::             |||      :::   |||::
   12801 TGAACGTGATACGCGTCAGCGTACGTGCCGCGGCACAACCTCTTTCCC 12850

      177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgG 193
          :::      ||||:::||          |||   ::::::::|||:::
   12851 CACCGCGACGGACGACGACGCCGAGGACGATGCGGGATCGAAAGTTCCACGC 12900

      194 n............ArgSerProLeuGlyGlyAspValIleHisAlaHis 209
                      |||      |||:::  :::   |||     :::
   12901 AGGGCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGG 12950

      210 euAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGly 226
          |||:::      :::|||          :::          |||||||||
   12951 AGGACTTTGAGCCCGACGCG.........CGAACCGGGCATTAGTCCCCGC 12997

      227 AlaIleHisTyrCysAlaSer.........ThrSerGluGlnSerTrpIh 246
                |||   :::|||:::         |||:::   |||:::   ::

   12992 CGCGCACACGTGGCGGCCGCCGACCTGGTAACGGCATACGAGCAGACGGT 13041

      236 rAspSerGluValAspSerSerCysSerGlyGlnPro.............. 248
          :::::::|||:::::::   :::   :||   :::
   13042 GAACCAGCAGATTAACTTTCAAAAAAGCTTTAACAACCACGTGGTACGC 13091

      249 ..................................IleHisLeuTrpGluPhe 254
                                           :::|||||||||:|||
   13092 TTGTGGCGCGCGAGGAGGTGGCTATAGGACTGATGCATCTGTGGGACTTT 13141

      255 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle..... 269
          ::::::   |||       :::||||:|||      :::
   13142 GTAAGCGCGCTGGAGCAAAACCCAAATAGCAAGCCGCCATAGGCGCAGCT 13191

      269 ..................................................... 269

   13192 GTTCCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 13241

      270 ..........................ArgTrpLeu.............Asn 278
                                   ||||.|||               |||
   13242 TGCTAAACATAGTAGAGCCCGAGGGCCGCTGGCTGCTGCATTTGATAAAC 13291

      274 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgIe 290
             :::::::|||          |||   |||   :::::::|||
   13292 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 13336

      290 uTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerA 307
                           :::      |.:::::::::::::   |||
   13337 ..............GACAAGGTGGCCGCCATCAACTATTCCATGCTT.... 13379

      307 rgSerIleArgGluTyrTyrLysLysGlyIleIleArgLysPro 321
          ||:::   :::::::||   :::   |||   :::   |||
   13370 ..AGCCTGGGCAAGTTTTACGCCGCAAGATATACCATACCCCT 13411

seq_name= [illegible] COMP seq-US-08-374-483-6

seq_documentation_block:
; Sequence 6, Application US/08374483
; Patent No. 5880102
;  GENERAL INFORMATION:
;    APPLICANT:  GEORGE, SAMUEL E.
;    APPLICANT:  BLAZING, MICHAEL A.
;    TITLE OF INVENTION:  ADENOVIRAL VECTOR SYSTEM
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD, 8TH FLOOR
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/374,483
;      FILING DATE:  17-JAN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WILSON, MARY J.
;      REGISTRATION NUMBER:  32,955
;      REFERENCE/DOCKET NUMBER:  1379-83
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 816-4000
;      TELEFAX:  (703) 816-4100
;      TELEX:  200797 NIXN UR
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  34382 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US 08 374 483 6


alignment scores:
            Quality:    93.50              Length:     366
              Ratio:  0.592                Gaps:        18
 Percent Similarity:  43.169   Percent Identity: 23.497

alignment block:
 US 09 126 945B 2 x US 08 374 483 6   ..

 Align seq 1/1  to: US 08 374 483 6  from: 1  to: 44382

       20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa  36
          |||           ||||||||||||   :::   ||||||   |||
   11851 CCCCGGTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 11892

       36 lGlyLeuGluArqArqAspTrpSerProSerProProAlaThrProGlu.  52
          ||||||      |||          |||                ||||||
   11893 .GGTTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAA 11941

       53 GluGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl  69
          ||||||   :::   |||                 :::|||:::|||::
   11942 CAGGGACGAGCCCCTTTT.................TTGCTTTCCCAGA 11978

       69 aAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArqGluGlu  85
          :   |||:::   ||||||   ||||||   :::|||::: |||:::|||
   11979 TGCATCCGGTGCTGCGGCAGATGCGCCCCCTCCTCAGCAGCGGCAAGAG 12028

       86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
              ::|||   :::|||                |||       ||:::
   12029 .....CAAGAGCAGCGCACACATGCAGGGCACCTCCCCTCCTCTAC 12067

      102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
          ::::|||                ||||||   |||
   12068 CGCGTCA..............GGAGGGGCGACA................. 12086

      119 euGluGlnValGlnSerMetValValGlyGluVal.............. 130
            :::   |||::::::   :::   |||::
   12087 ..CCCCCCTTCACGCGCCAGCACATCGTGATTACGAACCCCCGCGGGGT 12134

      131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIl 143
                      |||:::|||   :::      |||   :::::
   12135 CGGGCCGGCACTACCTGGACTTGGAGGAGGGCGAGGGCCTGGCCGCGCT 12184

      143 eThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrpLeuL 160
          :   |||                |||||:::::      :::      |
   12185 AGGAGCGCCC............TCTCCTGAGCGGTACCCAAGGTGCCAGC 12222

      160 euTrpThrGluHisGluTyrArqLeuProProMetGlyLysAlaPheGlu 176
          ||      :::   :::             |||      :::   ||:::
   12223 TGAAGCGTGATACGCGTGAGGCGTACGTGCCGCGGCAGAACCTGTTTCGC 12272

      177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluHePheArqEl 193
          :::     |||:::|||          |||  ::::::::::|||:::
   12273 GACCGCGAGGGAGAGGAGCCCGAGGAGATGCGGGATCGAAAGTTCCACGC 12322

      193 n............ArqSerProLeuGlyGlyAspValLeuHisAlaHisL 209
                     |||      |||:::   :::   |||      :::
   12323 AGGGCGCGAGCTGCGGCATGGCCTGAATCGCCAGCGGTTGCTGCGCGAGG 12372

      206 euAspIleTrpLysSerAlaAlaTrpMetLysGluArqThrSerProGly 222
          |||:::      :::|||          :::          |||||||:::
   12373 AGGACTTTGAGCCCGACGCG.........CGAACCGGGATTAGTCCCGCG 12413

      223 AlaIleHisTyrCysAlaSer.........ThrSerGluGluSerTrpTh 236
              |||   :::|||:::          |||:::   |||:::   ::
   12414 CGCGCACACGTGGCGGCGCGAGCTTGGTAACCGCATACGAGCAGACGGT 12463

      236 rAspSerGluValAspSerSerCysSerGlyGlnPro.............. 248
          ::::::|||::::::   :::   |||   :::


   12464 GAACCAGGAGATAAACTTCAAAAAAGCTTTAACAACCACGTGCGTACGC 12513

      249 ..............................IleHisLeuTrpGlnPhe 254
                                        :::|||||||||||:|||
   12514 TTGTGCCGCGCGAGGAGGTGCTATAGGACTGATGCATCTGTGGGACTTT 12563

      255 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArqPheHe..... 269
          ::::::   |||      :::|||:::|||         :::
   12564 GTAAGCGCGCTGGAGCAAAACCCAAATAGCAAGCCGCTCATGCGCAGCT 12613

      269 ..................................................... 269

   12614 GTTCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 12663

      270 ..........................ArqTrpLeu............Asn 278
                                     |||||||||         |||
   12664 TGCTAAACATAGTAGAGCCCGAGGGCCGCTGGCTGCTCGATTTGATAAAC 12713

      279 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArqLe 295
          :::::::|||          |||   |||   ::::::::|||
   12714 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 12758

      296 uTrpGlyIleArqLysAsnArqProAlaMetAsnTyrAspLysLeuSerA 312
                     :::      |||:::|||||||:::   |||
   12759 .............GACAAGGTGGCCGCCATCAACTATTCCATGCTT..... 12791

      313 rqSerIleArqGluTyrTyrLysLysGlyIleIleArqLysPro 421
          |||:::   ::::::|||   :::   |||   :::   |||
   12792 ..AGCCTGGGCAAGTTTACGCCCGCAAGATATACCATACCCCT 12833

Seq name: Seq. 2 is 12833 bp. Seq code: DNA. Seq is US 08 745 609 1

seq documentation_block:
; Sequence 1, Application 08/08745609
; Patent No. 5955360
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Hauser, Michael A.
;    APPLICANT:  Kumar-Singh, Rajendra
;    APPLICANT:  Hartigan O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States Of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/745,609
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Iniolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  BM 02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705 8410
;      TELEFAX:  (415) 397 8338
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35935 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
```

```
;      DESCRIPTION:  /desc = "DNA"
US-08-735-609-1


alignment_scores:
            Quality:   98.50              Length:     366
              Ratio:  0.592                Gaps:      18
 Percent Similarity:  43.169   Percent Identity: 28.497

alignment_block:
 US-09-126-945B-2 x US-08-735-609-1    ..

 Align_seq_1/1  to: US-08-735-609-1  from: 1  to: 45935

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 36
         |||         ||||||||||||    :::   ||||||   |||
   10928 CCCCGGTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 10969

      36 lGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu. 52
         ||||||      |||         |||                ||||||
   10970 .GGTTGCCTCCCTCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAA 11018

      53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl 69
         ||||||   :::   |||              :::|||:::|||::
   11019 CAGGGACGAGCCCCTTTT.................TTGCTTTTCCCAGA 11059

      69 uAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGlu 85
         :  ||||::   ||||||   ||||||   :::|||:: |||:::|||
   11151 TGCATCCGCTGCTGCGGCAGATGCGCCCCCCTCCTCAGCAGCGGCAAGAG 11100

      86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
         ::||||   :::|||              |||      ||||:
   11101 ......CAAGAGCAGCGGCAGACATGCAGGGCACCCTCCCCTCCTCCTAC 11144

     102 aGlySerLeuAspLeuValProGlyGlyLeuThrGluGluHisSerL 119
         :::|||                ||||||   |||
   11145 CGCGTCA..............GGAGGGGCGACA................ 11165

     119 euGluGlnValGlnSerMetValValGlyGluVal.............. 130
           :::   |||::::::   :::   |||:::
   11164 ..TCCGCGGTTGACGCGGCAGCAGATGGTGATTACGAACCCCCGCGGCGC 11211

     131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsnI 143
                     |||:::|||   :::      |||   :::::
   11212 CGGGCCCGCCACTACCTGGACTTGGAGGAGGGCGAGGGCCTGCCGCCGCCT 11261

     144 eThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuL 160
         :   |||               ||||||::::::      :::      |
   11262 AGGAGCGCCC............TCTCCTGAGCGGTACCCAAGGGTGCAGC 11299

     160 euTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGln 176
         ||      :::   :::              |||      :::   ||:::
   11300 TGAAGCGTGATACGCGTGAGGCGTACGTGCCGCGGCAGAACCTGTTTCGC 11349

     177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgSl 193
         :::      |||::::||          |||   ::::::::::||::::
   11350 GACCGCGAGGGAGAGGAGCCCGAGGAGATGCGGGATCGAAAGTTCCAGCC 11399

     193 n............ArgSerProLeuGlyGlyAspValLeuHisAlaHis 206
                     |||      ||||::   :::   |||      :::
   11400 AGGGCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGG 11449

     206 euAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGly 222
         |||:::      :::|||          :::          ||||||:::
   11450 AGGACTTTGAGCCCGACGCG.........CGAACCGCGATTACTCCCGCG 11490

     222 AlaIleHisTyrCysAlaSer.........ThrSerGluGlnSerTrpIl 237
           |||   ::||||:::          ||||::   |||:::   ::
   11491 CGCGCACACGTGGCGGCCGCCGACCTGGTAACCGCATACGAGCAGACGGT 11540

     238 rAspSerGluValAspSerSerCysSerGlyGlnPro............. 248
         :::::::|||:::::   :::   |||   :::
   11541 GAACCACGAGATTAACTTTCAAAAAAGCTTTAACAACCACGTGCGTACGC 11590

     249 ..............................IleHisLeuTrpGluPhe 254
                                        :::||||||||||:::|||
   11591 TTGTGGCGCGCGAGGAGGTGGCTATAGGACTGATGCATCTGTGGGACTTT 11640

     255 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle..... 269
         ::::::   |||       :::|||:::|||         :::
   11641 GTAAGCGCGCTGGAGCAAAACCCAAATAGCAAGCCGCTCATGGCGCAGCT 11690

     269 ................................................... 269

   11691 GTTCCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 11740

     270 ..........................ArgTrpLeu............Asn 278
                                    ||||||||||          |||
   11741 TGCTAAACATAGTAGAGCCCGAGGGCCGCTGGCTGCTCGATTGATAAAC 11790

     274 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLe 290
           ::::::||          |||   |||   :::::::|||
   11791 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 11835

     290 uTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerA 307
                         :::      |||:::||||||:::   |||
   11836 .............GACAAGGTGGCCGCCATCAACTATTCCATGCTT.... 11880

     307 rgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
           |||:::   :::::::|||   :::   |||   :::   |||
   11869 ..AGCCTGGGCAAGTTTTACGCCCGCAAGATATACCATACCCCT 11910

seq_name: [illegible] seq-US-08-735-609-1

seq_documentation_block:
; Sequence 1, Application US/08735609
; Patent No. 5994132
;   GENERAL INFORMATION:
;        APPLICANT: Chamberlain, Jeffrey S.
;                   Amalfitano, Andrea
;                   Hauser, Michael A.
;                   Kumar-Singh, Rajendra
;                   Hartigan-O'Connor, Dennis J.
;        TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
;        NUMBER OF SEQUENCES: 15
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Medlen & Carroll, LLP
;             STREET: 220 Montgomery Street, Suite 2200
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States of America
;             ZIP: 94104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/735,609
;             FILING DATE: 23-Oct 1996
;             CLASSIFICATION: Unknown
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Inqolia, Diane E.
;             REGISTRATION NUMBER: 40,027
;             REFERENCE/DOCKET NUMBER: UM-02484
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 705-8410
;             TELEFAX: (415) 397-8338
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 45935 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;             DESCRIPTION: /desc = "DNA"
```

```
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US 08 735 609 1


Alignment scores:
            Quality:   94.50              Length:     866
              Ratio:   0.592               Gaps:      18
 Percent Similarity:  43.169   Percent Identity: 23.497

Alignment block:
 US 09 126 945B 2 x US 08 735 609 1   ..

 Align seq 1/1  to: US 08 735 609 1  from: 1  to: 85945

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 36
         |||          |||||||||||||   :::   ||||||   |||
   10928 CCCCGCTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 10969

      46 tGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu 62
         ||||||      |||         |||                ||||||
   10970 .GGTTGCCTCCCCGCCATGCAAGACCCCGCTTGCAAATTCCTCCCGAAA 11018

      63 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl 69
         ||||||   :::   |||                 :::|||:::|||::
   11019 CAGGGACGAGCCCCTTTT.................TTGCTTTTCCCAGA 11060

      69 aAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGlu 85
         :   |||:::  ||||||  ||||||  :::|||::: |||:::|||
   11051 TGCATCCGGTGCTGCGGCAGATGCGCCCCCTCCTCAGCAGCGGCAAGAG 11100

      86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
         ::::||   :::|||                ||||      |||::
   11101 .....CAAGAGCAGCGGCAGACATCCAGGGCACCCTCCCCTCCTCCTAC 11144

     102 rGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
         ::::|||                |||||   |||
   11145 CGCGTCA.............GGAGGGCGACA................. 11183

     119 euGluGlnValGluSerMetValValGlyGluVal............... 130
           :::   |||::::::  :::   |||:::
   11184 ..TCCGCGGTTGACGCGGCAGCAGATGGTGATTACGAACCCCGCGGCGC 11211

     131 ............LeuLysAspIleGluThrAlaCysLysGluLeuAsnLe 143
                     |||:::|||   :::    |||   :::::
   11212 CGGGCCCGGCACTACCTGCACTTGGAGGAGGGCGAGGCCTGGCGCGGCT 11261

     143 eThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrpLeuL 160
         :   |||                |||||:::::::      :::      (
   11262 AGGAGCGCCC............TCTCCTGAGCGGTACCCAAGGGTGCAGC 11299

     160 euTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlu 176
         ||      :::   :::            |||       :::   |||:::
   11300 TGAAGCGTGATACGCGCGACGGCGTACGTGCCGCGGCAGAACCTGTTTCGC 11349

     177 GluLeuArgAlaGlyLysGluLeuCysAlaMetYsGluGluHisGlnLeuArgL 193
         :::       |||::|||          |||   :::::::::|||:::
   11350 GACCGCGAGGGAGAGGAGCCCGAGGAGATGCGGGATCGAAAGTTCCACGC 11399

     193 n............ArgSerProLeuGlyGlyAspValLeuHisAlaHisL 209
                    |||       |||:::   :::   |||       :::
   11400 AGGGCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGG 11449

     206 euAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGly 222
         |||:::       :::|||          :::          |||||:::
   11450 AGGACTTTGACGCCGACGCG.........CGAACGCGGATTAGTCCCGCG 11490

     223 AlaIleHisTyrCysAlaSer.........ThrSerGluIleSerTrpPh 239
               |||   :::|||:::          |||:::   |||:::   ::
   11491 CGCGCACACGTGGCGGCCGCGACCTGGTAACCGCATACGAGCAGACGGT 11540

     240 rAspSerGluValAspSerSerCysSerGlyGlnPro.............. 248
         :::::::|||::::::   :::   |||   :::
   11541 GAACCAGGAGATTAACTTTCAAAAAGCTTTAACAACCACGTGCGTACGC 11590

     249 ...............................IleHisLeuTrpGlnPhe 254
                                        :::|||||||||||:|||
   11591 TTGTGGCGCGCGAGGAGGTGGCTATAGGACTGATGCATCTGTGGGACTTT 11640

     255 LeuLysGluLeuLeuLysProHisSerTyrGlyArgPheIle..... 269
         ::::::   |||       :::|||:::|||        :::
   11641 GTAAGCGCGCTGGAGCAAAACGCAAATAGCAAGCCGCTCATCGCGCAGCT 11690

     269 ................................................... 269

   11691 GTTCCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 11740

     270 ............................ArgTripLeu.............Asn 278
                                      ||||||||||            |||
   11741 TGCTAAACATAGTAGAGCCCGAGGGCCGCTGGCTGCTCGATTGATAAAC 11790

     274 LysGluLysGlyIlePheLysIleGluAspSerAlaGluValAlaArgLe 290
         ::::::|||          |||   |||   :::::::|||
   11791 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 11835

     290 uTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerA 307
                         :::       |||:::|||||:::   |||
   11836 .............GACAAGGTGGCCGCCATCAACTATTCCATCCTT.... 11868

     307 rgSerIleArgGluTyrTyrLysLysGlyIleIleArgLysPro 321
         ||||::   :::::::|||   :::   |||   :::   |||
   11869 ..AGCCTGGGCAAGTTTTACGCCCGCAAGATATACCATACCCCT 11910

Search_Date: Tue Sep 18 17:29:45 2001
Search_Time: 1.5 seconds
Run_Time: 1.5 seconds
Seq_Name: US09126945B2.rni:US09126945B2.seq.ds on 09/452,483

seq_documentation_block:
; Sequence 43, Application US/08379452
; Patent No. 6040174
;  GENERAL INFORMATION:
;    APPLICANT:  IMLER, Jean Luc
;    APPLICANT:  MEHTALI, Majid
;    APPLICANT:  PAVIRANI, Andrea
;    TITLE OF INVENTION:  DEFECTIVE ADENOVIRUSES AND CORRESPONDING
;    TITLE OF INVENTION:  COMPLEMENTATION LINES
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  1737 King Street, Suite 500
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22314 2756
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/379,452
;      FILING DATE:  26 JAN 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR94/00624
;      FILING DATE:  27-MAY 1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 93 06482
;      FILING DATE:  28 MAY 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dadio, Susan M.
;      REGISTRATION NUMBER:  40,373
;      REFERENCE/DOCKET NUMBER:  029395 302
;  INFORMATION FOR SEQ ID NO:  43:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  45935 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;     MOLECULE TYPE:  DNA (genomic)
US-08-479-452-43


alignment_scores:
            Quality:   93.50              Length:     366
              Ratio:   0.592               Gaps:      18
 Percent Similarity:   43.169   Percent Identity: 23.497

alignment_block:
 US-09-126-945B-2 x US-08-479-452-43   ..

 Align seq 1/1   to: US-08-479-452-43  from: 1  to: 35945

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 46
         |||         |||||||||||||   :::   ||||||   |||
   10928 CCCGCGTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 10979

      36 tGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu. 52
         ||||||      |||         |||                ||||||
   10970 .GGTTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAA 11018

      53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl 69
         ||||||   :::   |||                 :::|||:::|||::
   11019 CAGGGACGAGCCCCTTTT.................TTGCTTTCCCAGA 11059

      69 nAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGlu 85
         :   ||||::   ||||||  ||||||  :::|||::: |||::|||
   11051 TGCATCCGGTGCTGCGGCAGATGCGCCCCCCCTCCTCAGCAGCGGCAAGAG 11100

      86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
         ::|||   ::|||          |||     ||||
   11101 ......CAACAGCATCGGCAGACATGCAGGGCACCCTCCCCTCCPCCTAC 11144

     102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGlnHisSerL 119
         :::|||               ||||||   |||
   11145 CGCGTCA..............GGAGGGGCGACA................. 11184

     119 euGluGlnValGlnSerMetValValGlyGluVal............... 130
         :::   |||||:::   :::   |||:::
   11164 ..TCCGCGTTGACGCGGCAGCAGATGGTGATTACGAACCCCGCGCGCGC 11213

     131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIl 148
                     ||||:|||   :::    |||   :::::
   11212 CGCGCCCGCCACTACCTGCACTTCGACGCAGGCCAGCTGCTGCAGGGT 11261

     143 eThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuL 160
         :   |||               ||||||:::::      :::     |
   11262 AGGACGCGCG............TCTCGTGAGCGGTACCCAAGGGTGCAGC 11299

     160 euTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGln 176
         ||    :::  :::          |||     :::   ||||:
   11300 TGAAGCGTGATACGCTGACGCCTACGTGCCGCCGCAGAACCTGTTTCGC 11349

     177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGl 193
         :::      |||:::|||         |||   :::::::::|||:::
   11350 GACCGCGAGGGAGAGGAGCCCGAGGAGATCGCGGATCGAAAGTTCCACGC 11399

     193 n............ArgSerProLeuGlyGlyAspValLeuHisAlaHis 206
                     |||      ||||::   :::   |||      :::
   11400 AGGCGCGAGCTGCGCCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGG 11449

     206 euAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGly 222
         ||:::      :::|||          :::          ||||||:::
   11450 AGGACTTTCAGCCCGACGCG.........CGAACCGGCATTAGTCCCGCG 11496

     223 AlaIleHisTyrCysAlaSer.........ThrSerGluGluSerIrpIh 238
             |||   :::|||:::          ||||::   ||||::   ::
   11491 CGCGCACACGTGGCCGCCGCCGACCTGGTAACCGCATACGAGCAGACGGT 11540

     239 rAspSerGluValAspSerSerCysSerGlyGlnPro.............. 248
         :::::::|||:::::   :::   |||   :::
   11541 GAACCAGGAGATTAACTTTCAAAAAAGCTTTAACAACCACGTGCGTACGC 11590

     249 ..............................IleHisLeuTrpGlnPhe 254
                                        :::|||||||||||:|||
   11591 TTGTGGCGCGCGAGGAGGTGGCTATAGGACTGATGCATCTCTGGGACTTT 11640

     255 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle..... 269
         ::::::   |||       :::|||:::|||          :::
   11641 GTAAGCGCGCTGGAGCAAAACCCAAATAGCAAGCGCTCATCGCGCAGCT 11690

     269 ..................................................... 269

   11691 GTTCCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 11740

     270 ............................ArgTrpLeu.............Asn 278
                                      |||||||||           |||
   11741 TGCTAAACATAGTAGAGCCCGAGGGCCGCTGGCTGCTCCATTTGATAAAC 11790

     274 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLe 290
         :::::::|||          |||   |||   ::::::|||
   11791 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 11835

     290 uTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerA 307
                           :::     |||:::|:: ::::   |||
   11836 ............GACAAGCTGGCCGCCATCAACTATTCCATGCTT.... 11868

     307 rnSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
           |||:::   :::::::|||   :::   |||   :::   |||
   11869 ..AGCCTGGGCAAGTTTTACGCCGCCAAGATAIACCATACCGCT 11910

seq_name: /cgn2_6/ptodata/1/iia/6A_COMB.seq_US-09-315-472-1

seq_documentation_block:
; Sequence 1, Application US/09315472
; Patent No. 6057158
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Hauser, Michael A.
;    APPLICANT:  Kumar-Singh, Rajendra
;    APPLICANT:  Hartigan-O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States Of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/315,472
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/735,609
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ingolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM-02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35935 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
```

```
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc   "DNA"
US-09-415-872-1


alignment_scores:
            Quality:   93.50              Length:     366
              Ratio:   0.592               Gaps:      18
 Percent Similarity:  43.109   Percent Identity:  23.497

alignment block:
 US-09-126-945B-2 x US-09-415-872-1   ..

 Align seq 1/1  to: US-09-415-872-1  from: 1  to: 35935

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 46
         |||          ||||||||||||   :::   ||||||   |||
   10928 CCCCCGTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 10969

      46 lGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu. 52
         ||||||      |||          |||                ||||||
   10970 .GGTTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAA 11018

      53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl 69
         ||||||   :::   |||                 ::||||:::|||:
   11019 CAGGGACGAGCCCCTTT.................TTGCTTTCCCAGA 11050

      69 aAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGlu 85
         :   ||||::   ||||||   ||||||   :::|||::: |||:::|||
   11051 TGCATCCGGTGCTGCGGCAGATGCGCCCCCCTCCTCAGCAGCGGCAAGAG 11100

      86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
         :::|||   :::|||                ||| |||::
   11101 .....CAAGAGCAGCGGCAGACATGCAGGGCACCCTCCCCTCCTCCTAC 11144

     102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
         :::|||           ||||||   |||
   11145 CGCGTCA...............GGAGGGGCGACA................. 11164

     119 euGluGlnValGlnSerMetValValGlyGluVal.............. 130
         :::   |||:::::::  :::   |||::
   11164 ..TCGCGCGTTGACGCGGCAGCAGATGGTGATTACGAACCCCCGCGGCGC 11211

     131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIl 143
                     |||:::|||   :::      |||   :::::
   11212 CGGGCCCGGCACTACCTGGACTTGGAGGAGGGCGAGGGCCTGGCGCGGCCT 11261

     143 eThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuL 160
         :   |||                ||||||:::::      :::      |
   11262 AGGAGCGCCC............TCTCCTGAGCGGTACCCAAGGGTGCAGC 11299

     160 euTrpThrGluHisGluTyrArgLeuProProMetGlyLysAlaPheGln 176
         ||     :::  :::             |||      :::   ||||::
   11300 TGAAGCGTGATACGCCTGAGGCGTACGTGCCGCGGCAGAACCTGTTTCGC 11349

     177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGl 193
         :::      |||:::|||       |||   ::::::::::::::::
   11350 GACCGCGAGGGAGAGGAGCCCCAGGAGATGCCGGATCGAAAGTTCCACGC 11399

     193 a.............ArgSerProLeuGlyGlyAspValLeuHisAlaHisL 206
                      |||      |||:::   :::   |||      :::
   11400 AGGCCGCGAGCTGCGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGC 11449

     206 euAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGly 222
         |||:::      :::|||          :::          |||||||:::
   11450 AGGACTTTGAGCCCGACGCG.........CGAACCGGGATTAGTCCCGCG 11490

     223 AlaIleHisTyrCysAlaSer.........ThrSerGluGluSerTrpTh 236
             |||   ::::|||:::         |||:::   |||:::   ::
   11491 CGCGCACACGTGGCGGCCGCCGACCTCGTAACCGCATACGAGCAGACGGT 11540

     236 rAspSerGluValAspSerSerCysSerGlyGlnPro............. 248

         ::::::::|||:::::   :::   |||   :::
   11541 GAACCAGGAGATTAACTTTCAAAAAAGCTTTAACAACCACGTGCGTACGC 11590

     249 .............................IleHisLeuTrpGluPhe 254
                                     :::|||||||||::|||
   11591 TTGTGGCGCGCGAGGAGGTGGCTATAGGACTGATGCATCTGTGGGACTTT 11640

     255 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle..... 269
         ::::::   |||      :::|||:::|||      :::
   11641 GTAAGCGCGCTCCGACAAAACCCAAATAGCAAGCGCGCTCATGCGCAGCT 11690

     269 ..................................................... 269

   11691 GTTCCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 11740

     270 ............................ArgTrpLeu.............Asn 278
                                     |||||||||             |||
   11741 TGCTAAACATAGTAGAGCCCGAGGGCGCGTGGCTGCTCGATTTGATAAAC 11790

     274 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLe 290
         ::::::|||           |||  |||   ::::::|||
   11791 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGACGCTGGCT..... 11835

     290 uTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerA 307
                      :::      |||:::|||||:::   |||
   11836 ..............GACAAGGTGGCGCGCATCAACTATTCCATGCTT.... 11868

     307 rgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
         |||:::   ::::::::|   :::   |||   :::   |||
   11869 ..AGCCTGGGCAAGTTTACGCCCGCAAGATATACCATACCCCT 11910

Seq_name: [illegible]

seq_documentation_block:
; Sequence 1, Application US/09244752
; Patent No. 6063622
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Hauser, Michael A.
;    APPLICANT:  Kumar Singh, Rajendra
;    APPLICANT:  Bartlam-O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/244,752
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/745,609
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ingolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM 02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705 8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35935 base pairs
```

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc   "DNA"
US-09-244-752-1


alignment_scores:
            Quality:   93.50              Length:     366
              Ratio:  0.592               Gaps:        18
 Percent Similarity:  43.169   Percent Identity: 23.497

alignment_block:
 US-09-126-945B-2 x US-09-244-752-1   ..

 Align seq 1/1  to: US-09-244-752-1  from: 1  to: 35935

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 36
         |||          ||||||||||||   :::   ||||||   |||
   10928 CCCCGGTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 10969

      36 lGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu. 52
          ||:||       |||          |||                 ||||||
   10970 .GGTTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAA 11018

      53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl 69
         ||||||   :::   |||                 :::||||:::|||:
   11019 CAGGGACGAGCCCCTTTT.................TTGCTTTTCCCAGA 11058

      69 nAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGln 85
         :  ||||::   ||||||   ||||||   :::|||||: ||||:|||
   11051 TGCATCGGGTGCTGCGGCAGATGCGCCCCCCTCCTCAGCAGCGGCAAGAG 11100

      86 ProProGluGluProGluGlnCysProValIleAspSerGluAlaProAl 102
            :::|||   :::|||                 |||      ||||:
   11101 ......CAAGAGCAGCGGCAGCATCCAGGGCAGCCTCGCCCTCCTCCTAC 11141

     102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
         :::|||                |||||   |||
   11145 CGCGICA.............GGAGGGGCGACA.............. 11183

     119 euGluGlnValGlnSerMetValValGlyGluVal............... 130
          :::   |||||||:   :::   |||||
   11184 ..TCCGCGGTTGACGCGGCAGCAGATGGTGATTACGAACCCCCGCGGCGC 11211

     131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsnI 143
                     |||:::|||   :::      |||   :::::
   11212 CCCGCCCGCCACTACCTCCACTTGCAGCAGCGCGAGGCCTCGGCGCGAGTT 11261

     143 eThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuL 159
         :  |||                |||||:::::      :::      |
   11262 AGGAGCGCCC............TCTCCTGAGCGGTACCCAAGGGTGCAGC 11299

     160 euTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheIle 176
         ||     :::   :::          |||      :::   ||||::
   11300 TGAAGCGTGATACGCGTGAGGCGTACGTGCCGCCGCAGAACCTGTTTCCT 11349

     177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgSl 193
         :::     |||:::|||          |||   ::::::::::|||:::
   11350 GACCGCGAGGGAGAGGAGCCCGACGACGATGCGGGATCGAAAGTTCCACGC 11399

     193 n............ArgSerProLeuGlyGlyAspValLeuHisAlaHist 206
                    |||      ||||::   :::   |||      :::
   11400 AGGGCGCGACACTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGG 11449

     206 euAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGly 222
           ||:::      :::|||          :::          |||||||::
   11450 AGGACTTTGAGCCCGACGCG.........CGAACCGGGATTAGTCCCGGG 11499

     223 AlaIleHisTyrCysAlaSer.........ThrSerGluGluSerTrpTh 236
            |||   :::|||:::           ||:::   ||:::   ::
   11491 CGCGCACACGTGGCGGCCGCGACCTGGTAACCGCATACGAGCAGACGCT 11540

     236 rAspSerGluValAspSerSerCysSerGlyGluPro.............. 248
         :::::::|||::::::   :::   |||   :::
   11541 GAACCAGGAGATTAACTTCAAAAAAGCTTTAACAACCACGTGCGTACGC 11590

     249 ..............................IleHisLeuTrpGlnPhe 254
                                      :::|||||||||||:||
   11591 TTGTGGCGGCGAGGAGGTGGCTATAGGACTGATGCATCTGTGGGACTTT 11640

     255 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle..... 269
         ::::::   |||      :::|||:::||          :::
   11641 GTAAGCGCGCTGGAGCAAAACCCAAATAGCAAGCCGCTCATGGCGCAGCT 11690

     269 ................................................... 269

   11691 GTTCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 11740

     270 ........................ArgTrpLeu............Asn 278
                                 |||||||||          |||
   11741 TGCTAAACATAGTAGAGCCCGAGGGCGGCTGGCTGCTCGATTGATAAAC 11790

     274 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLe 290
         ::::::|||          |||   |||   ::::::|||
   11791 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 11835

     290 uTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerA 307
                       :::      |||:::|||||:::   |||
   11836 .............GACAAGGTGGCCGCCATCAACTATTCCATGCTT.... 11868

     307 rgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
          |||:::   :::::|||  :::   |||   :::   |||
   11869 ..AGCCTGGGCAAGTTTTACGCCCGCAAGATATACCATACCCCT 11910

seq_documentation_block:
; Sequence 1, Application US/09245497
; Patent No. 6083750
;  GENERAL INFORMATION:
;    APPLICANT:  Chamberlain, Jeffrey S.
;    APPLICANT:  Amalfitano, Andrea
;    APPLICANT:  Hauser, Michael A.
;    APPLICANT:  Kumar-Singh, Rajendra
;    APPLICANT:  Hartigan-O'Connor, Dennis J.
;    TITLE OF INVENTION:  IMPROVED ADENOVIRUS VECTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/245,497
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/735,609
;      APPLICATION NUMBER:  (B) FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Inqolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  UM 02484
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  1:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  45935 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "DNA"
US-09-245-497-1


Alignment scores:
            Quality:   93.50              Length:     366
              Ratio:  0.592               Gaps:        18
 Percent Similarity:  48.169   Percent Identity:  24.497

Alignment Block:
 US-09-126-945B-2 x US-09-245-497-1  ..

 Align seq 1/1  to: US-09-245-497-1  from: 1  to: 45935

       20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa  86
          |||          ||||||||||||   :::   ||||||   |||
   10928 CCCCGGTTCGAGTCTCGGACCGCC...CGGACTGCGGCGAACGGG..... 10969

       86 lGlyLeuGluArqArqAspTrpSerProSerProProAlaThrProGlu.  52
          ||||||      |||         |||                ||||||
   10970 .GGTTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAA 11018

       53 GluGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl  69
          ||||||   :::   |||                  :::|||:::|||:
   11019 CAGGGACGAGCCCCTTTT................TTGCTTTCCCAGA 11050

       69 uAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArqGluGlu  85
          :   |||:::   ||||||   ||||||   :::|||::: |||:::|||
   11051 TGCATCCGGTGCTGCGGCAGATGCGCCCCCCTCCTCAGCAGCGGCAAGAG 11100

       86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl  102
             :::|||   :::|||                 |||       |||::
   11101 .....CAAGAGCAGCGGCAGACATGCAGGGCACCCTCCCCTCCTCCTAC 11144

      102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL  119
          :::::|||                |||||   |||
   11145 CGCGCA.....................GGAGGGCGACA.............. 11184

      119 euGluGluValGluSerMetValValGlyGluVal.............. 130
          :::   |||:::::::   :::   ||||::
   11184 ..TCCGCGCTTGACGCGCAGCAGATGGTGATTACGAACCCCCGCGCGC 11231

      131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsn11  143
                      |||:::|||   :::      |||   :::::
   11232 CGGGCCCGGCACTACTCGACTTGGAGGAGGGCGAGGGCTGGCGCGGCT 11281

      143 eThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrpLeuL  160
          :   |||                |||||:::::      :::       |
   11282 AGGAGCGCCC.............TCTCCTGAGCGGTACCCAAGGCTGCAGC 11329

      160 euTrpThrGluHisGlnTyrArqLeuProProMetGlyLysAlaPheIle 176
          ||      :::  :::              |||      :::   ||:::
   11330 TGAAGCGTGATACGCGTGAGGCGTACGTGCCGCGGCAGAACCTGTTTGCC 11379

      177 GluLeuAlaGlyLysGluLeuCysAlaMetSerIleGluGlnThrArqIl  193
          :::       |||::|||          |||   :::::::::|||:::
   11380 GACCGCGAGGGACAGGAGCCCGAGGACATGCGGGATCGAAAGTTCCACGC 11429

      194 n.............ArqSerProLeuGlyGlyAspValLeuHisAlaHisL  206
                     |||      |||:::   :::   |||      :::
   11400 AGGCCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGTTTGCTGCGCGAGG 11449

      206 euAspIleTrpLysSerAlaAlaTrpMetLysGluArqThrSerProGly  222
          |||:::     :::|||              :::          |||||||:::
   11450 AGGACTTTGAGCCCGACCGC.........CGAACCGGGATTAGTCCCGCG 11490

      223 AlaIleHisTyrCysAlaSer.........ThrSerGluGluSerTrpTh  236
                |||   :::|||:::          |||:::   |||:::   ::
   11491 CGCGCACACGTGGCGGCCGCGACCTGCTAACCGCATACGAGCAGACGGT 11540

      236 rAspSerGluValAspSerSerCysSerGlyGluPro.............. 248
          ::::::::|||::::::   :::   |||   :::
   11541 GAACCAGGAGATTAACTTCAAAAAAGCTTTAACAACCACGTGCGTACGC 11590

      249 ................................IleHisLeuTrpGluPhe 264
                                          :::||||||||||::|||
   11591 TTGTGGCGCGCGAGGAGCTGGCTATAGGACTGATGCATCTGTGGGACTTT 11640

      265 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArqPheIle..... 279
          ::::::   |||       :::||||::|||          :::
   11641 GTAAGCGCGCGCTGCAGAAAACCCAAATAGCAAGCTGCTCATGGCGCAGCT 11690

      279 ..................................................... 279

   11691 GTTCCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGC 11740

      280 .........................ArqTrpLeu.............Asn 293
                                  |||||||||          |||
   11741 TGCTAAACATAGTAGAGCCCGAGGGCCGCTGGCTGCTCGATTTGATAAAC 11790

      294 LysGluLysGlyIlePheLysThrGluAspSerAlaGluValAlaArqLe 310
          ::::::|||              |||   |||   ::::::|||
   11791 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 11835

      310 uTrpGlyIleArqLysAsnArqProAlaMetAsnTyrAspLysLeuSerA 327
                          :::      |||:::|||||:::   |||
   11836 ............GACAAGGTGGCCGCCATCAACTATTCCATGCTT.... 11868

      327 rqSerIleArqGluTyrTyrLysLysGlyIleIleArqLysPro 321
          |||:::   ::::::|||   :::   |||   :::   |||
   11869 ..AGCCTGGGCAAGTTTACGCCCGCAAGATATACCATACCCCT 11910

seq_data . . . . . /tmp/STN/dir_1_1_1_pA_COMB.seq US-09-409-670-43

seq_documentation_block:
; Sequence 43, Application US/09409670
; Patent No. 6133028
;  GENERAL INFORMATION:
;    APPLICANT:  IMLER, Jean-Luc
;    APPLICANT:  MEHTALI, Majid
;    APPLICANT:  PAVIRANI, Andrea
;    TITLE OF INVENTION:  DEFECTIVE ADENOVIRUSES AND CORRESPONDING
;    TITLE OF INVENTION:  COMPLEMENTATION LINES
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;      STREET:  1737 King Street, Suite 500
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22314-2756
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/409,670
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/608/379,452
;      FILING DATE:  26-JAN-1996
;      APPLICATION NUMBER:  WO PCT/FR94/00624
;      FILING DATE:  27-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 93 06482
;      FILING DATE:  28-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dadio, Susan M.
```

```
;       REGISTRATION NUMBER:  45,473
;       REFERENCE/DOCKET NUMBER:  029395 0302
;   INFORMATION FOR SEQ ID NO:  43:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  45985 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-09-409-670-43


alignment_scores:
            Quality:   93.50              Length:     366
              Ratio:  0.592                Gaps:      18
 Percent Similarity:  45.169   Percent Identity: 23.497

alignment_block:
 US-09-126-945B-2 x US-09-409-670-43  ..

 Align_seq 1/1  to: US-09-409-670-43  from: 1  to: 45985

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 66
         :||          ||:|||||||||   :::   ||||||   |||
   10928 CCCCGGTTCGAGTCTCGGACCGGC...CGGACTGCGGCGAACGGG..... 10969

      67 IGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu. 52
          ||||||      |||      |||             ||||||
   10970 .GGPTTGCCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGGAAA 11018

      53 GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl 69
         ||||||   :::   |||                 ::|||::|||||:
   11019 CAGGGACGAGCCCCTTTT.................TTGCTTTCCCAGA 11060

      69 uAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer.ArgGluGlu 85
         :   ||:::   ||||||   ||||||   :::|||:: |||:||||
   11061 TGCATCCGGTGCTGCGGCAGATGCGCCCCCTCCTCAGCAGCGGCAAGAG 11110

      86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
            :::|||   :::|||          |||      ||||
   11111 .....CAAGAGCAGCGGCAGACATGCAGGGCAGCCTCCCCTCCTCCTAC 11144

     102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
         :::|||               ||||||   |||
   11145 CGCGTCA..............GGAGGGCGACA.................. 11183

     119 euGluGlnValGlnSerMetValValGlyGluVal.............. 130
            :::   |||||||   :::   ||||:
   11184 ..TCCGCGGTTGACGCGGCAGCAGATGGTGATTACGAACCCCCGCGGCGC 11211

     131 ............LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIl 143
                     |||:::|||   :::    |||   :::::
   11212 CGGGCCCGGCACTACCTGGACTTGGAGGAGGGCGAGGGCCTGGCGCGGCT 11261

     143 eThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuL 160
         :  |||           ||||||:::::   :::      |
   11262 AGCACCGCGC............TCTCCTGAGCGGTACCTAAGGGTGCAGC 11299

     160 euTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGln 176
         ||     :::   :::          |||     :::   ||||::
   11300 TGAAGCGTGATACGCGTGAGGCGTACGTGCCGCGCCACAACCTGTTTCGC 11349

     177 GluLeuArgGlyLysGluLeuCysAlaMetSerGluGluGluPheArgGl 193
         :::     |||:::|||         |||   :::::::::|||:::
   11350 CACCGCCAGGGAGAGGAGCCCGAGGAGATGCGGGATCGAAAGTTCCACGC 11399

     193 n............ArgSerProLeuGlyGlyAspValLeuHisAlaHisL 206
                   |||      ||:::   :::   |||      :::
   11400 AGGGCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGC 11449

     207 euAspIleTrpLysSerAlaAlaTrpMetLysGlnArgIleSerProGly 222
         ||:::   :::|||         :::      ||||||:::
   11450 AGGACTTTGAGCCCGACGCG.....CGAACCGGGATTAGTCCCGCG 11490

     223 AlaIleHistyrCysAlaSer.....IleSerGluGlnSerTrpHi 236
             |||   :::|||:::         ||:::   |||:::   ::
   11491 CGCGCACACGTGGCGGCCGCCGACCTGGTAACCGCATACGAGCAGACGGT 11540

     236 rAspSerGluValAspSerSerCysSerGlyGlnPro..... 248
         :::::::|||:::::::   :::   |||   :::
   11541 GAACCAGGAGATTAACTTTCAAAAAAGCTTTAACAACCACGTGCGTACGC 11590

     249 .....IleHisLeuTrpGlnPhe 254
                                  :::|||||||||:::|||
   11591 TTGTGGCGCGCGAGGAGGTGGCTATAGGACTGATGCATCTGTGGGACTTT 11640

     255 LeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle..... 269
         ::::::   |||      :::|||:::|||         :::
   11641 GTAAGCGCGCTGGAGCAAAACCGAAATAGCAAGCGCGTCATGCGCGCAGCT 11690

     269 ..... 269

   11691 GTTCCTTATAGTGCAGCACAGCAGGGACAACGAAGCATTCAGGGATGCGC 11740

     270 .....ArgTrpLeu.....Asn 278
                              |||||||||          |||
   11741 TGCTAAACATAGTAGAGCCCGAGGGCCGCTGGCTGCTCCGATTTGATAAAC 11790

     274 LysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgIe 290
         :::::::::|||          |||   |||   ::::::|||
   11791 ATCCTGCAGAGCATAGTGGTGCAGGAGCGCAGCTTGAGCCTGGCT..... 11835

     290 uTrpGlyIleArgLysAsnArgProAlaMetAspTyrAspLysLeuSerA 307
                  :::          |||:::||||||:::   |||
   11836 .............GACAAGGTGGCCGCCATCAACTATTCCATGCTT..... 11868

     307 rgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro 321
         |||:::   :::::::||   :::   |||   :::   |||
   11869 ..AGCCTGGGCAAGTTTTACGGCCGCAACATATACCATACCGCT 11910

seq_name: /usr/prod/gpo/pct/US95-08354A-1/PCTUS-COMB.seq:PCT US95-08354A-1

seq_documentation_block:
; Sequence 1, Application PCT/US95/08354A
;  GENERAL INFORMATION:
;    APPLICANT:  Temple University - Of The
;    APPLICANT:  Commonwealth System of Higher Education
;    TITLE OF INVENTION:  JAK3 PROTEIN TYROSINE
;    TITLE OF INVENTION:  KINASE AND DNA ENCODING THE SAME
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seidel, Gonda, Lavorgna
;      ADDRESSEE:  & Monaco, P.C.
;      STREET:  Suite 1800, Two Penn Center
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  U.S.A.
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5o inch, 720 Kb
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/08354A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/272,368
;      FILING DATE:  8 July 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monaco, Daniel A.
;      REGISTRATION NUMBER:  30,480
;      REFERENCE/DOCKET NUMBER:  6056 203 PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-8383
```

```
;       TELEFAX:  (215) 568 5549
;  INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4016 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single stranded
;       TOPOLOGY:  linear
PCT US95 08354A 1


alignment scores:
            Quality:   93.00              Length:    292
              Ratio:  0.715               Gaps:       17
 Percent Similarity:  44.521   Percent Identity: 22.268

alignment block:
 US 09 126 945B 2 x PCT US95 08354A 1/rev   ..

 Align seq 1/1  to reverse of: PCT US95 08354A-1  from: 1  to: 4016

        5 SerProGlyLeuSerSerValSerProSerHis...LeuLeuLeuProPr  20
          |||  |||||||||||||||||||||||||||       ||||||||||||
     1229 AGCCAGGGGATTACAGACCTTGACCCAGCACAGCCTCCTCTTCTTCCTCC 1180

       20 o.........................AspThrValSerArgThrGlyL  28
          |                        |||  |||          ||||
     1179 TTCCCACCTCCACTTGTGGGGGGGGGGGGACGAGGTT......AGGGGAC 1136

       28 euGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSer  44
          ||||   :::   :::|||||        |||   :::         |||
     1135 TGAAGACAGGCCTGGTCGGGTCATGCCCACTGATCCAATCCATACTATCG 1086

       45 ProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSe  61
          ||||||           |||          ||||||::::::   |||
     1085 CCTTCTGTCCGTCCCCCGCCC.........CTATCTAGTCTCACCCTT... 1047

       61 rTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaP  78
                  ::::::   |||:::   |||   |||::::::::::   |
     1046 .........CTAGTCACGCCCCAGGTCAGTCCCTGGGGCAGCTCCATCC 1007

       78 roGlyAlaSerSerArgGluGluProProGluGluProGluGlnCysPro  94
          |||||   :::   :::      ||||||                ||||||
     1006 CCGGACCCACCTTTAAATTGACCCCGCCC............CCCTGCCCC 969

       95 ValIleAspSerGlnAlaProAla...GlySerLeuAspLeuValProGl  110
          |||:::::|||      |||:::   :::::::::   :::|||
      968 GTAGTGCAATCAGCCCGCCCGGGACTAGCCACTTTCCTGATTGTTCGGAC 919

      110 yGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValV  127
           |||:::      ::::::::|||:::      ::::::
      918 GGGCTTCTTTACTAACCGCCACGCCTCCCCGCGGACC.............. 882

      127 alGlyGluValLeuLysAspIleGlnThrAlaCys.....TyrLeu  139
                ::::|||||||      ::::::           ::::|||
      881 .....CCGCTTCTCAAGCCCAGCCCCAGCACCTCCTGGTCCCCGGAGCTC 837

      141 LeuAsnIleThrAlaAspPro.....................MetAspTr  150
          ::::::      ::||||                       ::::||
      836 CAGGAGATGCCGTTGTCCCCCGCCACACGCAGAAGCCCCGGCTCCTCCTG 787

      150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGluTyrA  167
          |:::|||::::::
      786 GGCGCCCGGAGC..................................... 774

      167 rgLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeu  183
                |||      |||:::            ||||||
      773 .......CCCACACGGAAGGTCTCGGTGGTGCCCCCTGGA......... 741

      184 CysAlaMetSerGluGluGlnPhe      ....ArgGlnArgSerPr  196
          |||:::   |||:::::::::::::
      740 TGTAGCCCCTCCAGGTCCAGATATACTTGGCCAAGAGCCGTAGCGGTCGG 691

      196 oLeuGlyGlyAsp.....................ValLeuHisAlaH  205
```

```
          |||:::|||||                          ||||||:::   :
      690 CCTGGCAGGCGACCACACCGCAGCGCCAAGACCACGGTCCTGCCGATGC 641

      205 isLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArg..... 218
          ::|||   :::   |||          |||:::   :::|||
      640 GCCTGCGTGTCACGAAGTTCTGGCCCTGGATCACATCCCGCAGGCTGGGC 591

      219 ..................................ThrSerProGlyAla11 224
                                           :::||||||||||
      590 GGCAGACAGGCTTTGTAACTGACCGTCTTCAGCAGCTCTCCTGGGCCC.. 543

      224 eHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluVal  240
                  :::   |||:::      |||:::   ||:::::::
      542 .........TGGGCCTGCTCACGAGCCATCTGGGCCAAGTCCAGCACGG 503

      241 ..........AspSerSerCysSer 245
                  :::|||   ||||||
      502 CCAGGCTCAGCAACTCTCCCTGCTCC 477

Seq_name: /seq2_1/pct_data/1_1nq/6A_1998/MB_seq/US-08-840-062-4

seq_documentation_block:
; Sequence 4, Application US/08840062
; Patent No. 6117977
;  GENERAL INFORMATION:
;    APPLICANT:  LASKY, LAURENCE A.
;    APPLICANT:  WU, KAI
;    TITLE OF INVENTION:  TYPE C LECTINS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/840,062
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  P1019R1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225 3216
;      TELEFAX:  415/952 9881
;      TELEX:  910/371 7168
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4771 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US 08 840 062 4


alignment scores:
            Quality:   93.00              Length:    261
              Ratio:  0.819               Gaps:       14
 Percent Similarity:  43.678   Percent Identity: 24.755

alignment block:
 US 09 126 945B 2 x US 08 840 062 4/rev   ..

 Align seq 1/1  to reverse of: US 08 840 062 4  from: 1  to: 4771

        6 ProGlyLeuSerSerValSerPro.........SerHisLeuLeuLeuPr  19
```

```
      ||||:::|||      :::::::|||          :::|||||||:::
   2104 CCAGCTCTTCTTGTCCTGCAGCCGCTCTGAGCTGAACACCTTATAGCAAT 2054

     19 oProAspIhrValSerArq..............IhrGlyLeuGluLysA 41
        ||:::      |||          ||||||   :::::::
   2054 ACCCCACTTCCCTGTCACACCCCCACCCCTCCCCACACCACCCCACCCCC 2004

     41 laAlaAlaGlyAlaValGlyLeuGlu...ArqArqAspIrpSer..... 44
        |:::   |||          :::   ||||||   ||||||
   2003 CTGGCCGTGGGAICTGGCCCCGGCAGCICCGGCGTCACTGGAGTGCCCA 1954

     45 .....ProSerProProAlaIhrProGluGlnGlyLeuSerAlaPheIy 59
             |||:::      |||   ||||||:::|||   ::::::::::
   1954 GCTCTGCCCGCAGATGTAGCGGGCCCGGAACGAGGT...ACAGTTCTICA 1907

     59 rLeuSerIyrPheAspMetLeuIyrProGluAspSerSerIrpAlaAlaL 76
        ::::                ||||::      :::   ||:::
   1906 CCIC.....................CCACAGCCCCAIGCGGCTGCCAGTG 1858

     76 ysAlaProGlyAlaSerSerArqGluGluPro.................. 86
                |||::              |||
   1857 GCCAGCGCCACGCAGCGCCACGGCTGTACCGGGGCTGTTCCCGGTTCCA 1808

     87 .....ProGluGluProGluGln............ 92
                        |||   ||||||||||:::
   1807 GTGGGTCTACATGACTTCATCCCCACTGAGCCAGAACAAGGAGCCGGTGC 1758

     93 .....CysProValIleA 97
                          |||||||||:
   1757 TGTTCAGGTTCTGCACGGCCGTCCAGAACTACTCGGCCTGCCACGTTCTAG 1708

     97 spSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThr 113
        ||:::   |||      |||   |||:::   |||      |||
   1707 ATGAGGCTGCTGACGAAGGCCTGCTCGAACCTGTTGGTGATGGTGACCAG 1658

    114 LeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
        |||:::              |||:::   :::::::::||||||:::::
   1657 CTGAGAGCCATGGTCAGTGCACAGGCGCCGGGCCTCACTGTAGGTCACTT 1608

    130 lLeuLysAspIleGluThrAla.CysLysLeuLeuAsnIleThrAlaAsp 149
        ::::   :::      :::|||      |||   ::::::
   1607 GGTCTTCTCCCAGCCAGTAGCAGGATGGGCTGTGCCACGTCCAACCCTTC 1558

    147 ProMetAspIrpSerProSerAsnValGlnLysIrpLeuLeuIrpThrGl 163
           :::   ||||||   |||:::::::      ||||||   |||
   1557 CGGCAGCCATGGTCC...TCCTCGGCGGCCCCCTGGCTCAGCTGG..... 1508

    163 uHisGlnTyrArqLeuProProMetGlyLysAlaPheGlnGluLeuAlaG 180
                                  ||||||   ::::::   |
   1507 ............................CCTGCCTTCTTGCAGATGGATG 1458

    180 lyLysGlu...LeuCysAlaMetSer.................GlnGln 199
        ||||:::   |||   :::::|||                   :::
   1457 GCAAGGACTGGTTACAGGGACTGTCGTTCCAGCGGCCTTCCGGGCCCCAG 1408

    199 GlnPheArqGlnArqSerProLeuGlyGlyAspValLeuHisAlaHisLe 209
        :::      |||   |||   |||:::      ::|||   :::
   1407 ATGGTGACACAGTCCTCCAGACTGTCCCGGAAGTTGTTGGGCTCA..... 1358

    209 uAspIleIrpLysSerAlaAlaIrpMetLys 216
                |||:::::   |||:::|||
   1357 ..........AAGGGCTGCCAGTGGGTGAAG 1328

Seq_name: ...

seq_documentation_block:
; Sequence 1, Application US/09263024
; Patent No. 6037159
; GENERAL INFORMATION:
;  APPLICANT: Uchimura, Kenji
;  APPLICANT: Muramatsu, Hideki
;  APPLICANT: Kadomatsu, Kenji
;  APPLICANT: Kannagi, Reiji
;  APPLICANT: Babuchi, Osami
;  APPLICANT: Muramatsu, Takashi
;  TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGALACTOSAMINE 6-O-SULFOTRANSFERASE AND
;  TITLE OF INVENTION: DNA ENCODING THE SAME
;  FILE REFERENCE: IOYAM41.001AUS
;  CURRENT APPLICATION NUMBER: US/09/263,023
;  CURRENT FILING DATE: 1999-03-05
;  EARLIER APPLICATION NUMBER: JP 10-54007
;  EARLIER FILING DATE: 1998-03-05
;  EARLIER APPLICATION NUMBER: JP 10-177844
;  EARLIER FILING DATE: 1998-06-24
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2150
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (470)...(1918)
US-09-263-023-1


alignment_scores:
            Quality:   92.50              Length:    250
              Ratio:    0.964               Gaps:     14
 Percent Similarity:   48.400   Percent Identity: 24.400

alignment_block:
 US-09-126-945B-2 x US-09-263-023-1   ..

 Align_Seq 1/1  to: US-09-263-023-1  from: 1  to: 2150

      22 IhrValSerArqThrGly................................ 27
         ||||||   ||:::|||
      93 ACAGTCCGCCCGGTCGGGGACATCTGGGGACGCTCCGAGGCGCACCCTCCGCT 142

      28 ..........LeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArqA 41
                   |||      ||||||   |||      :::|||
     143 CCAGGTCCTTCTCGGAGCCGCTGCCATGGTAGAGCCAGCCCTGGGCGCCG 192

      41 rqAspIrpSerProSerProProAlaIhrProGluGlnGlyLeuSerAla 57
         |||   :::|||   ||||||      |||:::   |||      :::
     193 GGGACCAGCAGCCTCTGCCGCCGCGCCCGCCTCGGATCGGCGGCCCCAGT 242

      58 PheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerIrpAl 74
                                     |||          |||
     243 ..............................CCCGGGCGCCGCAGCCGCCC 262

      74 aAlaLysAlaProGlyAlaSerSerArqGluGluProProGluGluProG 91
         |||   :::|||:::   ::::::          ||||||:::::
     263 TGCAGCGTCCCCCTCCTGGCCTGCAGGGCCGCCTCCGCCGCCGCCGCCC 312

      91 luGln.Cys.ProValIleAspSerGlnAlaProAlaGlySerLeuAspL 107
         ::: ||| ||||||:::::   :::::|||   :::          |
     313 CCCGGCTGTGCCCTGTGATGACGCCGCAGCTCGCCGGCGACGCT.....C 458

     107 euValProGlyGlyLeuIhrLeuGluGluHisSerLeuGluGlnValGlu 123
         ||   ||||||:::|||
     354 TGCCCCCCGGTGCGCCTT................................ 470

     124 SerMetValValGlyGluValLeuLysAspIleGluIhrAlaCysLysLe 140

     470 .................................................... 470

     140 uLeuAsnIleIhrAlaAspProMetAspIrpSerProSerAsnValGlnL 157
                         ||:::      :::|||:::   ||||||
     471 ............CCCCGGCTGCTGCCGGCCGCGCCTGCCGCCGTGCAGCC 407

     157 ysIrpLeuLeuIrpIhrGluHisGlnIyrArqLeuProProMetGlyLys 173
         ::   |||                    |||||||||
     408 GGGCCCTG.....................CTCCCGCCG......... 424
```

```
     174 AlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGl 190
     425 ...TGGCCCCGACCGCAGGAGCGCTGCTGGCCTGCG.............. 457

     190 nPheArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuA 207
     458 .............GCCCGCTGGG.....................AIGA 478

     207 spIleTrpLysSerAlaAla.TrpMetLysGluArgThrSerProGlyAl 223
     474 AGGTATTCGCAGGAAGGCGCTGGTGCTGTGGCGG.............GC 511

     223 aIleHis...TyrCysAlaSerThrSerGluGluSerTrpThrAspSerG 239
     512 TATGCACTGCTACTGGTGCTCACGATGCTCAACCTCTTGGACTACAACTG 561

     239 lnValAspSerSerCys...SerGlyGlnProIleHisLeuTrp 252
     562 GCATAAAGAGCCGCTGCAGCAGTGCAACCCCGACGGGCCTCTGG 605

seq name: /tmp/pbab/Dfsa/A.rMB...eq.fa sb 727 408 2

seq_documentation block:
; Sequence 2, Application US/08727408
; Patent No. 6020176
;  GENERAL INFORMATION:
;    APPLICANT:  Yuji KAMIYA et al.
;    TITLE OF INVENTION:  KAUKENE SYNTHASE
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/727,408
;      FILING DATE:  October 8, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202 371 8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2658 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US 08 727 408 2


alignment scores:
             Quality:    92.50              Length:     458
               Ratio:    0.578               Gaps:      17
 Percent Similarity:   44.693   Percent Identity: 21.508

alignment block:
 US 09 126 945B 2 x US 08 727 408 2  ..
```

```
Align seq 1/1  to: US 08 727 408 2  from: 1  to: 2658

       3 SerAlaSerPro.........GlyLeuSerSerValSerProSerHisLe 16
      75 AATGCCTCACCAGAGTTTGATGCTGCTTCCGCAATTTCTTTGCATGTCCT 124

      16 uLeuLeuProProAspThrValSerArq...ThrGlyLeuGluLysAlaA 32
     125 AAACATACCAGTCATGTATCTTTCCCGACCTACCGGCGTTGCCCGTTTTG 174

      32 laAlaGlyAlaValGlyLeuGluArqArqAspTrpSerProSerProPro 48
     175 CTGCCTCCTCCTCCTCCTCCTCCTCAGCTTCCCTATTCCCTGGAGTGGAT 224

      49 AlaThrProGluGlnGlyLeuSerAlaPheTyrLeu.............. 60
     225 GTGGACACAACAACTAAAACTGGAGCTTTCCACTTGAAGAAACAAAAGA 274

      61 ..........SerTyrPheAsp.............MetLeuTyrProGluA 70
     275 AAGAATTAAGAAATTGTTCGACAAGGTTGAACTTTCAGTTCTGCATATG 324

      78 spSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluPro 86
     325 ATACTGCATGGGTGGCAATGGTTCCTTCTCCAAACTCTCTCAACCAACCT 374

      87 ......ProGluGluProGluGluCysPro......ValIleAspSerGl 99
     375 CTTTGCCCGAG.............TGTATAAACTGGGTATTAGATAGTCA 412

      99 nAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluG 116
     413 ACATGCTGATGGCTCATGGGGCCTA.......................... 437

     116 luHisSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLys 132
     438 .......CTCCACAACGATCAGTTGCTGATGAAGGCCAATCTCTTATCT 479

     133 AspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAs 149
     480 ACATTA......GCATGTGTTCTT........................... 497

     149 pTrpSerProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnT 166
     498 .............ACTCTTAAGCGGTGGAATATTGGGCATGATCATA.... 530

     166 yrArqLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGlu 182
     531 .............ATGAGCAAGGCCCTTGATTTTATCAAGTCTAATATA 566

     183 leuCysAlaMetSerGluGluGlnPheArqGlnArqSerProLeuGlyGl 199
     567 GCTTCAGCTACTGATGAG.........AACCAACGTTCTCCGGTGGGATT 607

     199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetL 216
     608 TGACATTATTTTC.....................................620

     216 ysGluArqThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGlu 232
     621 .............CCTGGCATGATTGAGTATGCT.............. 641

     233 GluSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIl 249
     642 .........AAAGACTTGAATTTGAATTCTACCCTTGGCACCGACGAACGT 682

     249 eHisLeuTrpGlnPheLeuLysGluLeuLeuLeu.........LysProG 263
     683 CGATGCCTTGGTTCGAAAGAAAGAGTTGGAGCTGATGAGAAGCTGCAGAAGCA 732

     264 lsSerTyrGlyArqPheIleArqTrpLeuAsnLysGluLysGlyIlePhe 279
```

```
      733 ACTCTGAAGGTGGAAAAGCCTATTTAGCGTATGTTTCAGAAGGAATTGGA 782

      280 LysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAs 296
      783 AAGTTACAGGACTGGGATATGGTCATGCAATATCAA......AGGAAGAA 826

      296 nArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyr. 312
      827 TGGATCACTGTTTAATTCTCCATCCACTACGGCAGCGGCTTTTATGCATA 876

      313 .....TyrLysLysGlyIleIleArgLysPro 321
      877 GAAAIGATGATGGCTGTTTTGATTATCTTCGCTCACTGTTACAAAAGTTT 926

      322 AspIleSerGlnArgLeuValTyr 329
      927 GATGGCTCAGTTCCCACAATATAT 950

Seq_name: ????_? ?????? ? ??a FA COMP seq US-08-727-308-3

seq_documentation_block:
; Sequence 3, Application US/08727308
; Patent No. 6020176
;  GENERAL INFORMATION:
;    APPLICANT:  Yuji KAMIYA et al.
;    TITLE OF INVENTION:  KAURENE SYNTHASE
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/727,308
;      FILING DATE:  October 8, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2658 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-727-308-3


alignment_scores:
            Quality:   92.50          Length:    359
              Ratio:   0.578            Gaps:     17
 Percent Similarity:   44.693   Percent Identity:  21.508

alignment_block:
 US-09-126-945B-2 x US-08-727-308-3  ..

 Align seq 1/1  to: US-08-727-308-3  from: 1  to: 2658

      3 SerAlaSerPro.....GlyLeuSerSerValSerProSerHisLe 16
     75 AATGCCTCACCAGAGTTGATGCTGCTTCCGCAATTTCTTTGCATGTCCT 124

     16 uLeuLeuProProAspThrValSerArg...ThrGlyLeuGluLysAlaA 32
    125 AAACATACCAGTCATGTATCTTTCCGGACCTACCGGCGTTGCCCGTTTTG 174

     32 laAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProPro 48
    175 CTGCCTCCTCCTCCTCCTCCTCCTCAGCTTCCCTATTCCCTGGAGTGGAT 224

     49 AlaThrProGluGlnGlyLeuSerAlaPheTyrLeu..... 60
    225 CTGCAGCACAACAACTAAAACTGGAGCTTTGCACTTTCAAGAAACAAAAGA 274

     61 ..........SerTyrPheAsp.....MetLeuTyrProGluA 70
    275 AAGAATTAAGAAATTGTTCGACAAGGTTGAACTTTCAGTTTCTGCATATG 324

     70 spSerSerTrpAlaAlaLysAlaProGlyAlaSerArgGluGluPro 86
    325 ATACTGCATGGGTGGCAATGGTTCCTTCTCCAAACTCTCTCAACCAACCT 374

     87 ......ProGluGluProGluGlnCysPro......ValIleAspSerGl 99
    375 CTTTTCCCCGAG.....TGTATAAACTGGGTATTAGATAGTCA 412

     99 nAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluG 116
    413 ACATGCTGATGGCTCATGGGGCCTA..... 437

    116 luHisSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLys 132
    438 .....CTCCACAACGATCAGTTGCTGATGAAGGCCAATCTCTTATCT 479

    133 AspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAs 149
    480 ACATTA.....GCATGTGTTCTT..... 497

    149 pTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnT 166
    498 .....ACTCTTAAGCGCTGGAATATTGGGCATGATCAT.... 530

    166 yrArgLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGlu 182
    531 .....ATGAGCAAGGCCCTTGATTTTATCAAGTCTAATATA 566

    183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199
    567 GCTTCAGCTACTGATGAG.....AACCAACGTTCTCCGGTGGGATT 607

    199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetL 216
    608 TGACATTATTTTC..... 620

    216 ysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGly 232
    621 .....CCTGGCATGATTGAGTATGCT..... 641

    233 GluSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIl 249
    642 .....AAAGACTTGAATTTGAATCTACCCTTGGCACCGACGAACCT 682

    249 eHisLeuTrpGlnPheLeuLysGluLeuLeu.....LysProH 263
    683 GGATGCCTTGCTTCGAAAGAAACAGTTGGAGCTGAGAAGCTGCAGAAGCA 732

    263 isSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePhe 279
    733 ACTCTGAAGGTGGAAAAGCCTATTTAGCGTATGTTTCAGAAGGAATTGGA 782
```

```
      280 LysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAs 299
          |||:::::|||      :::|||   :::            ||||||||
      784 AAGTTACAGGACTGGGATATGGTCATGCAATATCAA.....AGGAACAA 826

      296 nArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyr. 312
          |         :::|||      :::   :::   :::      ::::::
      827 TGGATCACTGTTTAATTCTCCATCCACTACGGCAGCGGCTTTTATGCATA 876

      313 .....TyrLysLysGlyIleIleArgLysPro 321
                        |||   ||||||||||||||||
      877 GAAATGATGATGGCTGTTTTGATTATCTCGCTCACTCTTACAAAAGTTT 926

      322 AspIleSerGluArgLeuValTyr 329
          |||   |||         :::|||
      927 GATGGCTCAGTTCCCACAATATAT 950

Seq name: /cgn2_6/ptodata/1/iea/5A_COMB.seq:US_08_050_684-1

Seq documentation block:
; Sequence 1, Application US/08050684
; Patent No. 5550221
;  GENERAL INFORMATION:
;     APPLICANT:  Johann Dr., Stephen V.
;     APPLICANT:  Van Zeijl Dr., Marja
;     APPLICANT:  O'Hara Dr., Bryan M.
;     TITLE OF INVENTION:  Amphotropic Virus Receptor
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE:  American Cyanamid Company
;        STREET:  1937 West Main Street
;        CITY:  Stamford
;        STATE:  CT
;        COUNTRY:  United States of America
;        ZIP:  06904 0060
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE:  Floppy disk
;        COMPUTER:  IBM PC compatible
;        OPERATING SYSTEM:  PC DOS/MS DOS
;        SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER:  US/08/050,684
;        FILING DATE:  16 APR-1993
;        CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;        NAME:  Lowney Dr., Karen A
;        REGISTRATION NUMBER:  31,274
;        REFERENCE/DOCKET NUMBER:  31037 00
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE:  203 321 2361
;        TELEFAX:  203 321 2971
;        TELEX:  710 474 4059
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH:  4175 base pairs
;        TYPE:  nucleic acid
;        STRANDEDNESS:  single
;        TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;        NAME/KEY:  CDS
;        LOCATION:  244..2202
US 08 050 684 1


alignment scores:
             Quality:   92.50              Length:    317
               Ratio:  0.661               Gaps:       16
 Percent Similarity:  44.795   Percent Identity: 21.767

alignment block:
 US 09 126 945B 2 x US 08 050 684 1/rev  ..
```

```
Align Set 1/1  to reverse of: US 08 050 684 1  from: 1  to: 4175

        2 GlySerAlaSerProGlyLeuSerSer..... 10
          |||||||:::  |||||||||||:|||
     1375 GGTAGTTGCTTTCCTGGGCTGGCTTCTCCTCGGGGCCCCTGTCGATGTGG 1326

       11 .ValSerProSerHisLeuLeuLeuProProAspThrValSerArgThrG 27
                  :::   |||         ||||::       :::|||   |
     1325 ATTTTGTCGCAGCAGATCTTTGTAGAGCCCGAGTCTTTGTGCACGGTGTG 1276

       27 lyLeuGlulysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrp 43
          ||   :::                          :::|||||||:|||
     1275 GTACACAT.........................GACCGTCGCTCCTGG 1258

       44 Ser..ProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyr 59
             ||||||   |||   :::|||
     1252 TGTGGCCCTCGAAGCCGAAGGTGCCG..................... 1227

       60 LeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAla..... 74
          |||:::       |||:::      |||            ||||::
     1226 TTGGAGATGGGCGATTCACAGAGCCA.............TGGGTCATGGA 1189

       75 .AlaLysAlaProGlyAlaSerSerArgGluGluProProGluGluProG 91
          :::    |||  |||:::          |||  |||  |||:
     1188 CAGTGCTCTTCCGTATGCAGCCCGAGGTGGCTGCCCCGCAGAAGTGCCTT 1139

       91 lnGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeu 107
          :::::   |||  ::::::       ||||||::
     1138 CCGAGGTCCCCAGTGTCTCCCCTGCTGCTCCCGTG.............. 1104

      108 ValProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSe 124
             :::|||:::::|||:::::   :::|||         ::::|
     1103 ......AGCGGGATGGTGCTGTCATCATTAGCCTTGGCACCTGGTAGCTC 1060

      124 rMetValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuL 141
          |:::   :::|||:::                ::::::::      ||||
     1059 TTTAAATACTGGGGAC.............TCTGCTTCCTGAACCTTAC 1025

      141 euAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGluLys 157
          ||:::::::::::::|||      |||   :::||||||      :::
     1024 TGAGGCTTTCGTCAGATACTCGTGATAAAGCACCTTCTTTTGTAATTTG 975

      158 TrpLeuLeuTrpThrGlu...HisGlnTyrArgLeuProProMetGlyLy 173
          :::::::::   :::   |||   :::                    ||
      974 CCTGGTATTTTCCTCGGCATCCACGGACACACGAAGAGCCACACAAAAAA 925

      173 sAlaPheGlnGlnLeuAlaGlyLysGluLeuCysAlaMetSerGluGlnG 190
          ||||   ::::::   :::   ||||||||:   ||||||
      924 AGCGAACACGACGGGCGACACCAAAGGAAATGAGGGCTATG.......... 885

      190 lnPheArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu 206
                                                |||||||:::
      884 ..................................GCCCACATG 876

      207 AspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAl 223
                                          ||||||   |||
      875 GGG.............................AGAACAAGCCCGAG... 859

      223 aIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluV 240
          |||:::|||:::   :::   :::   :::
      858 ....CACTGGTGCTCCTGTGTACATGATGGAAAA.............. 829

      240 alAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLys 256
            |||   :::|||   ||||::   |||::   |||:::
      828 ..GACATTGATTCCTATGGTAGCAGCATAGAATACTGGGAGTGC..... 787

      257 GluLeuLeuLeuLysProHisSerTyrGly...ArgPheIleArgIrpLe 272
                        |||::::::   |||   |||   :::   :::
      786 ..............CCGGAGGCCATTGGGAACAGGGTCTCCTTTTTA 752

      272 nAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaA 289
          ::::::::::::      :::|||   :::        ::::::   :
```

```
      751 AGATGAAAATTCTGATGAGTACAAACAGCAGGCCAGACATGAAACCAGAC 702

      289 rnLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeu 305
          ::   |||      |||   :::      :::   |||   |||||||||
      701 AACAGTGG......AGATATAAACCAAGAAGCAACAATCTTGACAAGCTC 658

Seq_name: /...

Seq_documentation_block:
; Sequence 1, Application US/08582719
; Patent No. 5634348
; GENERAL INFORMATION:
;    APPLICANT: Johann Dr., Stephen V.
;    APPLICANT: Van Zeijl Dr., Marja
;    APPLICANT: O'Hara Dr., Bryan M.
;    TITLE OF INVENTION: Amphotropic Virus Receptor
;    NUMBER OF SEQUENCES: 2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: American Cyanamid Company
;      STREET: 1937 West Main Street
;      CITY: Stamford
;      STATE: CT
;      COUNTRY: United States of America
;      ZIP: 06904-0060
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/582,719
;      FILING DATE: 04-JAN 1996
;      CLASSIFICATION: 530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/050,684
;      FILING DATE: 16-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Lowney Dr., Karen A
;      REGISTRATION NUMBER: 31,274
;      REFERENCE/DOCKET NUMBER: 31937-00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 203-321-2461
;      TELEFAX: 203-321-2971
;      TELEX: 710-474-4059
;  INFORMATION FOR SEQ ID NO: 1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 4175 base pairs
;      TYPE: nucleic acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;    MOLECULE TYPE: cDNA
;    HYPOTHETICAL: NO
;    ANTI-SENSE: NO
;    FEATURE:
;      NAME/KEY: CDS
;      LOCATION: 244..2202
US-08-582-719-1


alignment_scores:
            Quality:   92.50            Length:    317
              Ratio:   0.651              Gaps:     16
 Percent Similarity:  44.795   Percent Identity: 21.767

alignment_block:
 US-09-126-945B-2 x US-08-582-719-1/rev  ..

 Align seq 1/1  1  reverse  of US-08-582-719-1  from 1  to 4175

        2 GlySerAlaSerProGlyLeuSerSer..................... 16
          |||||||:::  |||||||||||:::|||
     1475 GGTAGTTGCTTTCCTGGGCTGGCTTCTCCTCGGGGCCCCTGTCGATGTGC 1426

       11 .ValSerProSerHisLeuLeuLeuProProAspThrValSerArgThrG 27
             :::  |||       ||:::     :::|||  |
     1425 ATTTGTGCAGCA'ATCTTGTAGAGCCCCGAGTCTTGTGCACGGTGTC 1276

       27 lyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrp 43
          ||  :::                    :::|||||||:::|||
     1275 CTA'A'AT.....GACCGTCGCTCCTGG 1253

       44 Ser..ProSerProProAlaThrProGluGlnGlyLeuSerAlaPheTyr 59
             ||||||  :||  ::|||
     1252 TGTGGCCGTCGAAGCCGAAGGTGCCG..... 1227

       60 LeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAla..... 74
          |||:::     |||:::     |||          ||:::
     1226 TTGGAGATGGGCGATTTCACAGAGCCA.....TGGGTCATGGA 1189

       75 .AlaLysAlaProGlyAlaSerSerArgGluGluProProGluGluProG 91
          :::    |||  ||:::          |||  |||  |||:
     1188 CAGTGCTCTTCCGTATGCAGCCCGAGGGTGGCTGCCCGCAGAAGTGCCTT 1139

       91 luGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeu 107
          :::::  |||  ::::::      |||||||::
     1138 CCGAGGTCCCCACGTGTCTCCCCTGCTGCTCCCGTG..... 1104

      108 ValProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGluSe 124
             :::|||:::::|||:::::  :::|||          :::||
     1103 ......AGCGGGATGGTGCTGTCATCATTAGCCTTGGCACCTGGTAGCTC 1060

      124 rMetValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuL 141
          |:::  :::|||:::                ::::::::     ||||
     1059 TTTAAATACTGGGCAC                TCTGCTTCCTGAACCTTAC 1025

      141 euAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGluLys 157
          |||||||||||||||     |||  :::||||||  :::
     1024 TGAGGCTTTCGTCAGATACTCGTGATAAAGCAGCTTCTTTTTGTAATTTG 975

      158 TrpLeuLeuTrpThrGlu...HisGlnTyrArgLeuProProMetGlyLy 173
             :::::::  :::  |||  :::          ||
     974 CCTGTTATTTTCCTCCGCATCCACGGACACACGAAGAGCCACACAAAAAA 925

      173 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluG 190
          ||||  :::::  :::  |||||||:  ||||||
     924 AGCGAACAGGAGGGCGACACCAAAGGAAATGAGGGCTATG..... 885

      190 luPheArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu 206
                                                |||||||::
     884 .....GCCCACATG 876

      207 AspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAl 223
                                          ||||||  |||
     875 GGG.....AGAACAAGGCCGAG... 859

      223 alleHisTyrCysAlaSerIhrSerGluGluSerIrpIhrAspSerGluV 240
             |||:::|||:::  :::  :::  :::
     858 ...CACTGGATCCTCTGTCACATGATGTAAAA..... 829

      240 alAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLys 256
          |||  :::|||  |||:::  |||:::  |||:::
     828 ..GACATTGATTGCTATGGTAGCAGCATAGAATACTGGGAGTGC..... 787

      257 GluLeuLeuGluLysProHisSerTyrGly...ArgPheIleArgIrpLe 272
                      |||:::::  |||  |||  :::  :::
     786 .............CCGGAGGCCATTGGCAACAGAGTCTTCCTTTTTA 752

      272 nAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaA 289
          ::::::::::::      :::|||  :::          ::::::   :
     751 AGATGAAAATTCTGATGAGTACAAACAGCAGGCCAGACATGAAACCAGAC 702

      289 rnLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeu 305
          ::   |||      |||   :::      :::   |||   |||||||||
      701 AACAGTGG......AGATATAAACCAAGAAGCAACAATCTTGACAAGCTC 658

seq_name: /...
```

```
Seq_documentation_block:
; Sequence 8, Application US/09144246B
; Patent No. 6207377
; GENERAL INFORMATION:
;  APPLICANT: Wayne, Jay
;  APPLICANT: Xu, Shuang yong
;  TITLE OF INVENTION: Method For Construction Of Thermus E. coli Shuttle
;  TITLE OF INVENTION: Vectors And Identification of Two Thermus Plasmid
;  TITLE OF INVENTION: Replication Origins
;  FILE REFERENCE: Thermus Shuttle Vector
;  CURRENT APPLICATION NUMBER: US/09/144,246B
;  CURRENT FILING DATE: 1998-08-14
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 11958
;   TYPE: DNA
;   ORGANISM: Thermus sp.
US-09-144-246-8


Alignment_scores:
             Quality:   92.50              Length:    217
               Ratio:    0.954               Gaps:     11
 Percent Similarity:   44.700   Percent Identity: 25.346

Alignment_block:
 US-09-126-945B-2 x US-09-144-246-8/rev  ..

 Align_seq 1/1  to reverse of: US-09-144-246-8  from: 1  to: 11958

       13 ProSerHisLeuLeuLeuProProAspThrValSerArgThrGlyLeuGl 29
          |||      ||||||||||  |||          |||      |||
     9840 CCCCTCGCCCTCCTTCTCCAACCA.........AGCCCGCACGGG..... 9805

       29 nLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrp..... 43
                               ||||||||||||      |||
     9804 ......................GGGTTGGAGAGGATAACCTGGATAGGGG 9777

       44 .....SerProSerProProAlaThrPro...GlnGlnGlyLeuSerAla 57
             |||  |||      |||:|||   :::::|||  ||||::
     9776 CCTTCAGCTGCTCGTTCAGGGCTCAGCGTATTCTTCCCGGAAGAGCGGG 9727

       58 PheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAl 74
          :::      ::::::::::   ||||||
     9726 ATGATGCCAAACTTTTTCTCG...CTATAC..................... 9700

       74 aAlaLysAlaProGlyAlaSerSerArgGluGluProProGluGluProG 91
                  |||::::::::::||||||::::::||||||||||:::|||
     9699 .........CCCAGCTCCGCCAGCCGAAGGAGTCCGCCAGAACGCCCCC 9660

       91 lnGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeu 107
          :::      |||:::   :::|||:::   ::::||   |||   :::
     9659 TTGAAGGGCACGTACTCCCCGGTCAGGGCCAGGGTGGTGTTCTCCACGTT 9610

      108 ValProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSe 124
             |||      ::::::   ||||||   :::|||
     9609 GGCCCTGAGGACGTAGTAGGGGAGGAGAAGGATCTC.............. 9574

      124 rMetValValGlyGlnValLeuLysAspIleGluThrAlaCysLysLeuL 141
             ||||||      |||      ::::::   |||          :
     9573 ......GTTGGCCCAGATCTCCCCCCGCTCCAGCTTGCCCTTGACCGCG 9531

      141 euAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGluLys 157
          ::                              :::|||:::
     9530 TC.............................GGCCCGGC............ 9520

      158 TrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAl 174
                                             |||:::|||
     9519 .....................................CCTTTCGGCTACCC 9506

      174 aPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnP 191
                |||:::   :::   :::      ||||||   :::      |||
     9505 GCTGCAGGATCCGGGTGACGAAGCTGCCCTGTGCCGGCGAAGGGCTCCAGG 9456

      191 heArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAsp 207
             ::::::      |||:::      :::
     9455 AIGAAGACCTICTCCCCATCGAGCCCCCGCCC.................. 9424

      208 IleTrpLysSerAlaAlaTrp...MetLysGluArgThrSerProGlyAl 223
             :::||||||:::|||      :::   |||:::||||||||||::
     9423 ......AACAGTGCTTCCTGGCCAGCTCGTCTGCGCTCTTCACCAGGAAG 9380

      223 a 223
            :
     9379 t 9379

Seq_name: /tmp/ramdisk/US-09-126-945B-2_RNI.seq.txt on 764.233A 1

seq_documentation_block:
; Sequence 1, Application: US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Nolt, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  15006/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541 8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49477 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1, pUL3, and pVKM15
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  383..760
;      OTHER INFORMATION:  /product= "SorR"
```

```
;       OTHER INFORMATION:  /note  "This gene encodes a protein that is highly homologous
;       OTHER INFORMATION:  the reductase domains of type I PKSs such as oryA from
;       OTHER INFORMATION:  Saccharopolyspora erythraea."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  927..19874
;       OTHER INFORMATION:  /product  "SorA"
;       OTHER INFORMATION:  /note  "Gene product is highly homologous to type I PKSs that
;       OTHER INFORMATION:  are known to be involved in the synthesis of polyketide
;       OTHER INFORMATION:  compounds."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  942..7115
;       OTHER INFORMATION:  /product  "Module 1 of sorA"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7203..12884
;       OTHER INFORMATION:  /product  "Module 2 of sorA"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  13455..19616
;       OTHER INFORMATION:  /product  "Module 3 of SorA"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19871..46418
;       OTHER INFORMATION:  /product  "SorB"
;       OTHER INFORMATION:  /note  "Gene product is highly homologous to type I PKS genes
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19870..24556
;       OTHER INFORMATION:  /product  "Module 1 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  24638..30820
;       OTHER INFORMATION:  /product  "Module 2 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  30881..35446
;       OTHER INFORMATION:  /product  "Module 3 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  35529..40214
;       OTHER INFORMATION:  /product  "Module 4 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  40190..46418
;       OTHER INFORMATION:  /product  "Module 5 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  46851..47891
;       OTHER INFORMATION:  /product  "SorM"
;       OTHER INFORMATION:  /note  "The protein encoded by the sorM gene is highly
;       OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;       OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;       OTHER INFORMATION:  polyketide rappamicin."
US-08-764-238A-1


alignment_scores:
            Quality:   92.50              Length:    292
              Ratio:   0.670                Gaps:     15
 Percent Similarity:  47.260   Percent Identity:  24.315

alignment_block:
 US-09-126-945B-2 x US-08-764-238A-1   ..

 Align seq 1/1  to: US-08-764-238A-1  from: 1  to: 49377

        4 AlaSerProGlyLeuSerSerValSer.....ProSerHisLeuLeuLe 18
          :::::::::::::      :::   :::      ::::::      :::::
     5758 ICTICACCCGCGGCGCCGTCTCCATCGGACACTCCGACCCCATCACTCAT 5807

       18 uProPro.....AspThrValSerArgT 26
```

```
          :::::::                                  ::::::::::::   :
     5808 CCCACCCAGGCCATGATCTGGGGCCTCGGCCGCGTCGTCGGCCTCGAGCA 5857

       26 hrGlyLeuGluLysAla.....AlaAlaGlyAlaValGlyLeuGlnArq 40
          ||       :::   :::      ::::::::::::::::      :::
     5858 CCCCGAGCGATGGGCGGGCTCGTCGACGTCAGCCTGGGGTCGACGAGA 5907

       41 ArgAspTrpSer.....ProSerProProAlaThrProGlu... 52
             ::::::            ::::::::   ::::::   :::
     5908 GCGCCGTGGGCAGCTTGCTCCCGCCTCGCCCAGCGCCACGACGAAGAC 5957

       53 .GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProG 69
        ::::::::::::::::
     5958 CAGCTCGCTCTCCGC.....CGGC 5977

       69 luAspSerSerTrpAlaArgLysAlaProGlyAlaSerSerArgGluGlu 85
          ::::::::   ::::::::::          ::::   :::
     5978 CGGACTCTACGTCGCCGCATCGTCCGTCCTCGCCGATCGCGCTC 6027

       86 Pro.....ProGluProGluGlnCysProValIleAs 97
          :::            ::::::   :::::::   :::
     6028 CCGCGCGGGAGTTAGACCCCGAGCCACCATCCTCATCACCG..... 6069

       97 pSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrL 114
             ::::::::::::                  ::::::::::::::::
     6070 ....GAGCACCGCG.....CGCTCGGCGCTCACGT 6097

      114 euGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluVal 130
             :::   ::::::   ::::::::::::
     6098 CGCCCGATGGCTCGCTCGCCAGGGCGCAGAGC..... 6129

      131 LeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspPr 147
             :::   ::::::::::::::::      :::      :::   ::
     6130 ...ACCTCATCCTCATCAGCCGCCGAGGCGCCGAGGCCCCTGGCG...CC 6178

      147 oMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluH 164
          |   :::   :::::::::::::::
     6179 TCGGAGCTCCACGCCGAGCTCAATGCCC..... 6209

      164 isGlnTyrArgLeuProProMetGlyLysAlaPheGlnGluLeuAlaGly 180
             :::      ::::::      ::::::::::::   :::
     6210 ..TCGCCGTTCGCACACGGTTCGCGCTTCGATCTCGCCGATA..... 6248

      181 LysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLe 197
          :::   :::      :::::::::      ::::::::   :::
     6249 GAAGCGCTCTCCAAGCTCTCCTCGACAGCATTCCGTCGGACTGCCCGCTC 6298

      197 uGlyGlyAspValLeuHisAlaHisLeuAspIleTrplysSerAlaAla1 214
                        :::   :::::                  :::::
     6299 ACGGCGGTGTTTCACACGGCAGGAGCTCGCGACGATGGCCTGATCGGCGA 6348

      214 rpMetLysGluArgThrSer.....ProGlyAlaIleHis 225
             :::      :::::                ::::::::::::
     6349 CATGACGCCCGAGCGCATCGAGCGGGTCCTTGCGCCCAAGCTCGATCGG 6398

      226 TyrCysAla.....SerThrSerGluGl 238
             ::::::                  ::::::   :::::
     6399 CGTTGCACTTGCACGAGCTCACGAAAAATAGCGCTCTCGACGCCTTCGTC 6448

      238 uSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleH 250
          :::          ::::::::::::::::::          :::::::::
     6449 CTCTACGCTTCACTCTCGGGTGTCCTCGGCAATCCCGGTCAGGCCAATTA 6498

      250 isLeuTrpGlnPheLeuLysGluLeu 258
             :::   ::::::::::
     6499 CGCCGCTGCAAACGCTTTCCTCGATG 6525

Seq name: Seq2 Sqp:4 Ap:1 Fr:+1 To:ZA  5'-MR -- pat US-08-046 585 15

seq_documentation_block:
; Sequence 15, Application US/08046585
; Patent No. 5453362
;  GENERAL INFORMATION:
```

```
;     APPLICANT:  Lamarco, Kelly
;     APPLICANT:  Wilson, Angus
;     APPLICANT:  Berr, Winship
;     TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN
;     TITLE OF INVENTION:  HOST CELL FACTOR
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;       STREET:  4 Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94111 4187
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/046,585
;       FILING DATE:  12 APR 1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Osman, Richard A
;       REGISTRATION NUMBER:  36,627
;       REFERENCE/DOCKET NUMBER:  A-57594-1/RA
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-3249
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8252 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-046-585-15


alignment scores:
            Quality:   92.00            Length:    401
              Ratio:   0.697              Gaps:     13
 Percent Similarity:  43.854   Percent Identity: 22.924

alignment_block:
 US-09-126-945B-2 x US-08-046-585-15  ..

 Align_seq 1/1  to: US-08-046-585-15  from: 1  to: 8252

       6 ProGlyLeuSerSerValSerProSerHisLeu..................  16
         ||||||   |||          |||||||||||
    4824 CCGGGCCCCTCTGTGCCCCCCCCAGAGGAACTCCAGGTGTCGCCAGGTCC 4873

      17 .......LeuLeuProProAspThrValSerArgThrGlyLeuGluLysA  41
               |||||||||          |||       |||||||||
    4874 TGGCCAGCAGCTGCCGCCA.............CGGCAGCTTCTGCAGTCGG 4911

      41 laAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPro  47
         ||||||||||   ||||||   |||   |||       |||   |||
    4912 CTTCCACAGCCCTGATGGGGGAGTCCGCCGAGGTCCTGTCAGCCTCC... 4958

      48 ProAlaThrProGluGluGlyLeuSerAlaPheTyrLeuSerTyrPheAs  64
               |||||||||          ||||||          |||       ||
    4959 ...CAGACCCCTGAGCTCCCGGCCGCCGTGGATCTGACCAGCACAGGGGA 5005

      64 pMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaS  81
         |            ||||||||||||   |||||   |||       |||
    5006 GCCATCTTCGGGCCAGGAGTCTGCCGGCTCTGCGGTGGTGGCCACTGTGG 5055

      81 erSerArgGluGluProProGluGluProGluGlnCysProValIleAsp  97
               |||       ||||||   |||   |||                |||||
    5056 TGGTCCAGCCACCCCCACCCACACAGTCCGAA.............GTAGAC 5093

      98 ..........................SerGluAlaProAlaGlySerIle 105
                                   ||||||||||   ||||||||||
    5094 CAGTTATCACTTCCCAAGAGTAATGGCCGAGCCCAACTGGCACCAC 5143

     105 uAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeu..... 119
           ||||||       |||||||||   ||||||   ||||||
    5144 CACCCTCATGGTAACGGGGCTCACCCCCGAGGAGCTGGCAGTGACGGCTG 5193

     120 ..............................GluGlnValGlnSerMet 125
                                      ||||||||||||||||||
    5194 CTGCAGAAGCAGCTGCCCAGGCCGCAGCCACGGAGGAAGCCCAGGCCCTG 5243

     126 ValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAs 142
         ||||||       ||||||          ||||||||   |||    |||
    5244 GCCATCCAGGCGGTGCTC.........CAGGCCGCGCAGCAGGCCGTCAT 5284

     142 nIleThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrpL 159
         ||||||||||||||||||   |||    |||    |||
    5285 GGGCACCGGCAGACCATGGACACCTCCGAGGCAGCAGCA..... 5324

     159 euLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPhe 175
               |||       |||   ||||||       ||||||
    5325 .......ACCCTGACTCAGGCGAGCTGGGCCACCTGTCGCCCGAGGGT 5366

     176 GlnGlu..........................LeuAlaGlyLysGluLe 183
         ||||||                        ||||||    ||||||||
    5367 CAGGAGGGCCAGGCCACCACCATACCCATTGTGCTGACACAGCAGGAGCT 5416

     183 uCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyA 200
         ||||||||    |||||||||          |||||||||          ||||
    5417 GGCTGCCCTGGTGCAGCAGCAGCAGCTGCAGGAGGCCCAG......GCCC 5460

     200 spValLeuHisAlaHisLeuAsp........................... 207
         ||    |||   ||||||
    5461 AGCAGCAGCATCACCACCTCCCCACTGAGGCCCTGGCCCCTGCCGACAGT 5510

     208 ................IleTrpLysSerAlaAlaTrpMetLysGluArgTh 219
                         |||   |||||||||                  ||
    5511 CTCAACGACCCAGCCATTGAGAGCAATTGCCTCAATGAGCCTGGCCGGCAC 5560

     219 rSerProGlyAlaIleHistyrCysAlaSerThrSerGluGluSer.Trp 235
         |   ||||||||||          ||||||||::   ||||||.|||
    5561 GGTCCCCAGCACTGTGGCGCTGCTGCCTCAACGGCCACTGAGAGCCTGG 5610

     236 ThrAspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTr 252
                                  ||||||   ||||||||
    5611 CTC.............................CATCCAACACATTGTG 5630

     252 p 252
         |
    5631 G 5641

seq_name: /usr2/pfdata2/files/5A_COMB.seq:US-08-493-703-15

seq_documentation_block:
; Sequence 15, Application US/08493703
; Patent No. 5585249
;  GENERAL INFORMATION:
;     APPLICANT:  Lamarco, Kelly
;     APPLICANT:  Wilson, Angus
;     APPLICANT:  Berr, Winship
;     TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN
;     TITLE OF INVENTION:  HOST CELL FACTOR
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;       STREET:  4 Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94111 4187
;     COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/393,703
;       FILING DATE:  24-FEB-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Osman, Richard A
;       REGISTRATION NUMBER:  36,627
;       REFERENCE/DOCKET NUMBER:  A-5750-2/RAO
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-4249
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8252 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-393-703-15


alignment_scores:
            Quality:   92.400              Length:     801
              Ratio:   0.697                Gaps:      13
 Percent Similarity:  43.854   Percent Identity: 22.924

alignment_block:
 US-09-126-945B-2 x US-08-393-703-15   ..

 Align seq 1/1  to: US-08-393-703-15  from: 1  to: 8252

        6 ProGlyLeuSerSerValSerProSerHisLeu..................  16
          ||||||   |||          ||||||||||||
     4824 CCGGGCCCCTCTGTGCCGCCCCCAGAGGAACTCCAGGTGTCGCCAGGTCC 4875

       17 .......LeuLeuProProAspThrValSerArgThrGlyLeuGluLysA  41
                 |||||||||          |||      |||||||||
     4874 TCGCCAGCAGCTGCCGCCA............CGGCAGCTTCTGCAGTCGG 4911

       41 laAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPro  47
          ||||||||||   ||||||   :::   :::       |||   |||
     4912 CTTCCACAGCCCTGATGGGGGAGTCCGCCGAGGTCCTGTCAGCCTCC... 4958

       48 ProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAs  64
             |||||||||          ::::::       |||      ::
     4959 ...CAGACCCCTGAGCTCCCGGCCCCCGTGGATCTGAGCAGCACAGGGCA 5005

       64 pMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaS  81
          :          ::::::|||:::   ::::||   :::      :::
     5006 GCCATCTTCGGGCCAGGAGTCTGCCGGCTCTGCGGTGGTCGCCACTGTCG 5055

       81 erSerArgGluGluProProGluGluProGluGlnCysProValIleAsp  97
             :::      ||||||   :::   |||              :::|||
     5056 TGGTCCAGCCACCCCCACCCACACAGTCCGAA............GTAGAC 5093

       98 ..........................SerGluAlaProAlaGlySerIle  105
                                   ::::::|||   |||||||:::
     5094 CAGTIATCACTTCCCCAAGAGCTAATGGCCGAGGCCCAAGCTGGCACCAC 5143

      105 aAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeu.......  119
          ||:::      ||||||||||   ||||||   ::::::
     5144 CAGCCTCATGGTAACGGGGCTCACCCCCGAGGAGCTCGCAGTGACGCTG 5193

      120 ..............................GluGlnValGlnSerMet  125
                                       ||||||||||||||||||
     5194 CTGCAGAAGCAGCTGCCCAGGCCGCAGCCACGGAGGAAGCCCAGGCCCTG 5243

      126 ValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAs  142
          :::::      ||||||          :::::|||   :::   :::
     5244 GGCATCCAGGGGATGCTC.........CAGGCCGCCA AGAAGGCGTCAT 5284

      142 nIleThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrpL  159
             |||::::|||||||||   |||   :::   :::
     5285 GGGCACCGGCCAGCCCATGCACACCTCCGAGGCAGCAGCA.......... 5424

      159 euLeuTrpIhrGlaHisGluTyrArqLeuProProMetGlyLysAlaPhe  175
             |||       |||   ::::||       ::::::
     5425 ........ACCGTGACTCAGGCGGAGCTGGGCCACCTGTCGGCCGAGGCT 5460

      176 CisGla.....................                IeuAlaGlyLysGlnLe  188
          ||||||                    ||||::   :::||||
     5461 CAGGAGGGCCAGGCCACCACCATACCCATTGTGCTGACACAGCAGGAGCT 5510

      189 uCysAlaMetSerGlnGlnGlnPheArgGluArgSerProLeuGlyGlyA  200
          |:::|||::   ::::::|||      |||::::::          ::::
     5417 GGCTGCCCTGGTGCAGCAGCAGCAGCTGCAGGAGGCCCAG.....GCCC 5460

      200 spValLeuHisAlaHisLeuAsp............................  207
          ::       |||   ||||||
     5461 AGCAGCAGCATCACCACCTCCCACTGAGGCCCTGGCCCCTGCCGACAGT 5510

      208 ...............IleTrpLysSerAlaAlaTrpMetLysGluArgTh  219
                         |||   :::::::::                  ||
     5511 CTCAACGACCCAGCCATTGAGAGCAATTGCCTCAATGAGCTGGCCGGCAC 5560

      219 rSerProGlyAlaIleHisTyrCysAlaSerIhrSerGluGluSerArg  235
          |   |||:::::::::          ||||||:::   ||||| |||
     5561 GGTCCCCAGCACTGTGGCGCTGCTGCCCTCAACGGCCACTGAGAGCCTGG 5610

      236 ThrAspSerGluValAspSerSerCysSerGlyGluProIleHisLeuPr  252
                                                 ::::||   ||||||||
     5611 CTC.............................CATCCAACACATTTGTG 5650

      252 p 252
             |
     5651 G 5651

seq_name: /tmp2_fgp: Data/US_Data_Files/US09_0009/seq_list_US09_0009_127_21-15

seq_documentation_block:
; Sequence 15, Application PCT/US9311721
;   GENERAL INFORMATION:
;     APPLICANT:  Lamarco, Kelly
;     APPLICANT:  Wilson, Angus
;     APPLICANT:  Berr, Winship
;     TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
;     TITLE OF INVENTION:  HOST CELL FACTOR
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;       STREET:  4 Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94111-4187
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US9311721
;       FILING DATE:  03-DEC-1993
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Osman, Richard A
;       REGISTRATION NUMBER:  36,627
;       REFERENCE/DOCKET NUMBER:  A-5750-2/RAO
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-4249
```

```
;       TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  8252 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT US93 11721 15


Alignment scores:
            Quality:   92.00            Length:    301
              Ratio:  0.697              Gaps:     13
 Percent Similarity:  44.854   Percent Identity:  22.924

Alignment block:
 US-09-126-945B-2 x PCT US93 11721 15  ..

 Align seq 1/1  to: PCT US93 11721 15  from: 1  to: 8252

       6 ProGlyLeuSerSerValSerProSerHisLeu..................  16
         ||||||   |||          |||||||||||
    4824 CCGGGCCCCTCTGTGCCCCCCCAGAGGAACTCCAGGTGTCGCCAGGTCC 4873

      17 .......LeuLeuProProAspThrValSerArgThrGlyLeuGluLysA  41
                |||||||||          |||       |||||||||
    4874 TCGCCAGCAGCTGCCGCCA.............CGGCAGCTTCTGCAGTCGG 4911

      42 laAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPro  47
         ||||||||||   ||||||   :::   :::        |||   |||
    4912 CTTCCACACCCCTGATGGGGGAGTCCGCCGAGGTCCTGTCACCCCC... 4958

      48 ProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAs  64
         |||||||||          ::::::          |||       ::
    4959 ...CAGACCCCTGAGCCTCCCGGCCGCCTGCATCTGAGCAGCACAGGGGA 5005

      64 pMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaS  81
         :          ::::::|||:::   :::|||   :::       :::
    5006 GCCATCTTCGGGCCAGGAGTCTGCCGGCTCTGCCGGTGTTGCCACTGTGC 5055

      81 erSerArgGlnGluProProGluGluProGlnGlnCysProValIleAsp  97
            :::       ||||||   :::   |||              :::|||
    5056 TGGTCCAGCCACCCCCACCCACACAGTCCGAA............GTAGAC 5093

      98 ............................SerGluAlaProAlaGlySerLe 105
                                     ::::::|||   ||||||:::
    5094 CAGTTATCACTTCCCCAAGAGCTAATGGCCGAGCCCAAGCTGGCACCAC 5143

     106 nAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeu....... 119
         |||:::       ||||||||||   ||||||   ::::::
    5144 CACCCTCATGGTAACGGGGCTCACCCCCGAGCAGCTGGCAGTGACGGCTG 5193

     120 ..............................GluGluValGlnSerMet  125
                                      ||||||:|||||||:::
    5194 CTGCAGAAGCAGCTGCCCAGGCCGCAGCCACGGAGGAAGCCCAGGCCTG 5243

     126 ValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuA: 112
         ::::::      ||||||          :::::|||   :::   :::
    5244 GCCATCCAGGCGGTGCTC.........CAGGCCGCGCAGCAGGCCGTCAT 5284

     142 nIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpL 159
         |||:::::|||||||||    |||   :::   :::
    5285 GGGCACCGGCGAGCCCATGGACACCTCCGAGGCAGCAGCA.......... 5324

     159 euLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPhe 175
             |||       |||   ::|||       ::::::
    5325 ........ACCGTGACTCAGGCGGAGCTGGGGCACCTGTCGGCCCAGGGT 5366

     176 GlnGlu.............................LeuAlaGlyLysGluLe 188
         ||||||                           |||:::   :::||||||
    5367 CAGGAGGGCCAGGCCACCACCATACCCATTGTGCTGACACAGCAGGAGCT 5416

     189 uCysAlaMetSerGluGluGluPheArgGluArgSerProLeuGlyGlyA 208
         |:::|||:::   :::::::|||      |||::::::          ::::
    5417 GGCTGCCCTGGTGCAGCAGCAGCTGCAGGAGCCCCAG......GCCC 5460

     209 spValLeuHisAlaHisLeuAsp............................ 217
           ::    |||   ||||||
    5461 AGCAGCAGCATCACCACTCCCCACTGAGGCCCTGGCCCTGCCCGACAGT 5510

     218 ...............IleTrpLysSerAlaAlaTrpMetLysGluArgTh 219
                        |||   :::::::::                  ||
    5511 CTCAACGACCCAGCCATTGAGAGCAATTGCCTCAATGAGCTGGCCGGCAC 5560

     219 rSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSer.Trp 235
         |   ||:::::::::            ||||:::::   ||||||.|||
    5561 GGTCCCAGCACTGTGGCCCTGCTGCCCTCAACGGCCACTGAGAGCCTGG 5610

     236 ThrAspSerGluValAspSerSerCysSerGlyGluProIleHisLeuTr 252
                                      ::::||   ||||||||
    5611 CTC.............................CATCCAACACATTGTG 5650

     252 p 252
         |
    5651 G 5651

..g_nam=./2/2/2-1.seq Ty="G" Len=5651 Check=6454 MBase=2.03 on 228.162.4.1

seq documentation block:
; Sequence 1, Application US/08228162
; Patent No. 5521071
;  GENERAL INFORMATION:
;     APPLICANT:  Attie, Alan D
;     APPLICANT:  Sturley, Stephen L
;     APPLICANT:  Gretch, Daniel G
;     TITLE OF INVENTION:  Soluble LDL Receptor and Gene
;     NUMBER OF SEQUENCES:  2
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Nicholas J. Seay, Quarles & Brady
;       STREET:  P.O. Box 2113
;       CITY:  Madison
;       STATE:  WI
;       COUNTRY:  USA
;       ZIP:  53701 2113
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  08/08/228,162
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/925,403
;       FILING DATE:  03 AUG 1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seay, Nicholas J
;       REGISTRATION NUMBER:  27,386
;       REFERENCE/DOCKET NUMBER:  96269.2026
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (608)251-5000
;       TELEFAX:  (608)251-9166
;  INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1417 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     IMMEDIATE SOURCE:
```

```
;       CLONE:  LDL Receptor
;     FEATURE:
;        NAME/KEY:  CDS
;        LOCATION:  16..1086
US-08-228-162-1


alignment_scores:
            Quality:   91.50              Length:     392
              Ratio:  0.602               Gaps:        17
 Percent Similarity:  38.776   Percent Identity: 19.898

alignment_block:
 US-09-126-945B-2 x US-08-228-162-1    ..

 Align_seq 1/1  to: US-08-228-162-1  from: 1  to: 1417

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeu.. 16
          |||||||||||||||   :||   ||||||||||:::   :::
      368 ATGGGAAGTGCATCTCTCGGCAGTTCGTCTGTGACTCAGACCGGGACTGC 417

       17 .............LeuLeuProProAspThrValSerArgThrGlyLeuG 29
                       ||||||          |||   :::
      418 TTGGACGGCTCAGACGAGGCCTCCTGCCCGGTGCTCACCTGTGGTCCCGC 467

       29 luLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerPro 45
          :::::::||   |||
      468 CAGCTTCCAGTGCAACAGCT.............................. 487

       46 SerProProAlaThrProGlnGlnGlyLeuSerAlaPheTyrLeuSerTy 62
          ||||.||||::::::::::   |||
      488 ...CCACCTGCATCCCCCAGCTGTGGG..................... 511

       62 rPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProG 79
                     |||     ::::::   :::||||||
      512 .............CCTGCGACAACGACCCCGACTGCGAAGATGGCTC 545

       79 lyAlaSerSerArgGluGln.............................. 85
          ||:::::|||::
      546 GGATGAGTGGCCGCAGCGCTGTAGGGGTCTTTACGTGTTCCAAGGGCACA 595

       86 ...ProProGluGluProGluGlnCysPro......ValIleAspSerGl 99
             |||   :::|||:::::                |||   ::::::::
      596 GTAGCCCCTGCTCCGGCCTTCGAGTTCCACTGCCTAAGTGGCCAGTGCATC 645

       99 nAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluG 116
          :::|||||||||:::::   :::::|||   :::   |||
      646 CACTCCAGCTGGCGCTCTGATCGTGGCCCCGACTGCAAGGACA          688

      116 luHisSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLys 132

      688 ..................................................... 688

      133 AspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAs 149
                                 |||:::|||
      689 .........................AATCTGACG.............. 697

      149 pTrpSerProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnI 166
                           :::|||       ||||||
      698 ...................AGGAAAACTGCGCTGTGG.............. 715

      166 yrArgLeuProProMet.............GlyLysAlaTyrGluIleLeu 182
                |||||||::                :::::|||:::::::|||
      716 ........CCACCTGTCGCCCTGACGAATTCCAGTGCTCTGATGGAAACT 757

      182 AlaGlyLysGluLeuCysAlaMetSerGluGluGlnIleArgGln..... 199
          ||:::          :::::::::   ::::::      :::
      758 GCATCCATGGCAGCCGGCAGTGTGACCGGGAATATGACTGCAAGGACATG 807

      194 ..................................ArgSerProLeuG 198
                                           |||:::|||   :
      808 AGCGATCAAGTTGGCTGCGTTAATGTGACACTCTGCGAGGGACCCAACAA 857

      198 lyGlyAspValLeuHisAlaHisLeuAspIleTrp...LysSerAlaAla 213
          :::::::::|||      |||:::   :::   |||   ||||||||||:::
      858 GTTCAAGTGTCACAAGGGGAATGCATCACCCTGGACAAAGTCTGCAACA 907

      214 IrpMetLys.............................................. 216
          |||::::::
      908 TGGCTAGAGACTGCCGGGACTGGTCAGATGAACCCATCAAAGAGTGCGGG 957

      217 ...................................................GluArgI 219
                                                            |
      958 ACCAACGAATGCTTGGACAACAACGGCGGCTGTTCCCACGTCTGCAATGA 1007

      219 hrSerProGlyAlaIleHistyrCysAlaSerThrSerGluGluSerIrp 235
          ||      ::::||          ||||||   |||:::::::::   |||
     1008 CCTTAAGATCGGCTACGAGTGCCTGTGCCCCGACGGCTTCCAGCTGGTGG 1057

      236 ThrAsp.................SerGluValAspSerSerCysSerGl 246
             :::                 ||||::::::||          |||||
     1058 CCCAGCGAAGATGCGAAGATCTACTCTAGACTAGATCGATGAGTGTCAGG 1107

      246 yGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProH 263
          ||::                 :::   |||:::    |||    :::|||
     1108 ATCC...............CGACACCTGCAGCCAGCTCTGCGTGAACCTG 1142

      263 isSerTyrGlyArgPheIleArgIrpLeuAsnLysGluLysGlyIlePhe 279
          :::    :::                 :::   |||:::|||   |||
     1143 GAGGGTGCTACAAG............TGCCAGTGTGAGGAAGGC...TTC 1177

      280 LysIleGluAspSerAlaGluValAlaArgLeuIrpGlyIleArg..... 294
          ::::::      :::   :::   |||||||||||||:::|||
     1178 CAGCTGACCCCCACACGAAGGCCTGCAAGGCTGTGGTCCATCGCCTACCT 1227

      295 .......LysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerI 309
                 ||||||          ::::::          |||:::||||||
     1228 CTTCTTCACCAACCGGCACGAAGGTCAGGAAGATGACGCTGGACGGGACGS 1277

      309 leArgGlnTyrTyrLysLysGlyIle 317
          ::::::   :::    |||:::
     1278 AGTACGAATTCCTGCAGCCGGGGTC 1303

seq_name: /cgn2_6/ptodata/1/pira/PA_*.MP_seq_us-09-059-369_1

seq_documentation_block:
; Sequence 1, Application US/09059369
; Patent No. 6040156
;  GENERAL INFORMATION:
;    APPLICANT:  KAWASAKI, TOSHISUKE
;    APPLICANT:  OKA, SHOGO
;    TITLE OF INVENTION:  DNA ENCODING GLUCURONYLTRANSFERASE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/059,369
;      FILING DATE:  14-APR-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 9-127065
;      FILING DATE:  16-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
```

```
;       REFERENCE/Deposit Number:  ATCC 98662
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703 413 3000
;       TELEFAX:  703 413 2220
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2090 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     ORIGINAL SOURCE:
;       ORGANISM:  WISTAR RAT
;       TISSUE TYPE:  CEREBRAL CORTEX
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  195..1235
;       OTHER INFORMATION:  /note= "DEFINE METHOD: P"
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  252..302
;       OTHER INFORMATION:  /note= "TRANSMEMBRANE DOMAIN,
;       OTHER INFORMATION:  DEFINE METHOD"
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  651..659
;       OTHER INFORMATION:  /note= "POTENTIAL N-GLYCOSYLATION
;       OTHER INFORMATION:  SITE, DEFINE METHOD: S"
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  784..791
;       OTHER INFORMATION:  /note= "POTENTIAL GLYCOSYLATION
;       OTHER INFORMATION:  SITE, DEFINE METHOD: S"
;     FEATURE:
;       NAME/KEY:  misc feature
;       LOCATION:  1140..1148
;       OTHER INFORMATION:  /note= "POTENTIAL N-GLYCOSYLATION
;       OTHER INFORMATION:  SITE, DEFINE METHOD: S"
US-09-059-469-1


Alignment scores:
            Quality:   91.50              Length:    345
              Ratio:   0.683               Gaps:     17
 Percent Similarity:  42.029   Percent Identity: 19.420

Alignment block:
 US-09-126-945B-2 x US-09-059-469-1   ..

 Align seq 1/1  to: US-09-059-469-1  from: 1  to: 2090

        4 AlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProPr  20
          :::||||||||||       |||:::::|||:::|||
       19 AGCAGCCCTGGG......BCTCTGGGGCCAGGGCAT.............  48

       20 oAspThrValSerArgThrGlyLeuGluLysAla.............AlaA  33
             |||:::::|||   |||   :::::::::              ||||
       49 .AGGACTGCCACCCGCTATGGACCGCCCAGGGACGATATGGACTCGCTG  97

       34 laGlyAlaValGlyLeuGluArg.........ArgAspTrpSerProSer  46
          |:::::::::::::|||:::||                 ||:::   |||
       98 CCGCAGGTATCAACCTCCGAAGGTTCCTGACCCTGCGCTGGACTACCTCC  147

       47 ProProAlaThrProGluGluGlyLeuSerAlaPheTyrLeuSerTyrPh  63
          |||       ||||||   :::       ||||||      |||       ::
      148 CCTTCGCAGACTCCCATCAGGCCGACTCTGCAAACCTGCTCCCACAATG  197

       64 oAspMet...LeuTyrProGluAspSerSerTrpAlaAlaLysAlaProG  79
          :   |||            :::::::::   |||   ::::::      |
      198 GGTAATGACGACGCTGTGGCCGCAGCCAGCCCTTGGACGATGCCGAAGACGAAG  247

       80 lyAlaSerSerArgGluGluProProGluGluProGluGlnCysProVal  95
          |:::|||||||||   :::          |||   :::|||
```

```
      248 GGACATCCTCGCGATTGTCCTCATTGTGCTTCCCTGGACACTGC.....  291

       96 IleAspSerGluAlaProAlaGlySer.....................  104
                :::::::|||::|||:::
      292 .........TCATCACCGTCTGGCACCAGAGCAGCCCTCGCACCTCTGCTT  332

      105 ................LeuAspLeuValProGlyGlyLeuThrLeuGluG  116
                                :::::::::|||          ::::::::::
      333 GCTGTGCACAAGGATGAGGGAAGTGACCCCC.........GCCATGAGGC  373

      116 luHisSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLys  132
          ::|||   ::::::   :::   |||
      374 ACCACCCGGTGCGGACCCTACGGAGT.....................  399

      133 AspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAs  149
                |||||||||
      400 .........ACTGCATGT.............................  408

      149 pTrpSerProSerAsnValGluLysTrp..LeuLeuTrpThrGluHisGl  165
                     ::|||   :::|||  ::::::   |||||||:::
      409 .........CCGACCGTGACATCGTGGAGGTGGTGCCACACAGCTACGT  448

      165 nTyrArgLeuProPro..............................  170
           |||       ||||||
      449 GTACAGGAGGCCGCCACCGTGGTCCGACACGCTGCCCACCATCCATGTGG  498

      171 ...........MetGlyLysAlaArgSerArgGluLeuAlaGlyLysGluLeu  183
                        ::::::   ||:::   ||||||
      499 TGACGCCCACCTACAGTAGACCGGTGCAGAAG.....GCAGAGCTG  539

      184 CysAlaMetSerGlu..............................  188
               |||:::::::
      540 ACGCGAATGGCAACACGCTATTGCATGTCCCAACCTTCACTGGCTGGT  589

      189 ....GluGlnPheArgGluArgSerProLeu..............  197
             |||:::      :::|||:::||||||
      590 GGTGGAGGATGCTCACGTAGGACGCCCTCACAGCGCGCTGCTGCGG  639

      198 .....GlyGlyAspValLeuHisAlaHisLeuAspIle.......  208
               |||   :::      |||   |||:::::::
      640 ACACTGCCTCAACTATACACACCTGCACGTAGAGACACCACGCAACTAC  689

      208 ...........................................  208

      690 AAGCTGCGAGGTGACGCCGAGACCCTCGCATCCCACGTGGCACCATGCA  739

      209 ..................IrpLysSerAlaAlaTrpMetLysGluArgI  219
                           |||      ::::::   ::::::   |
      740 GCGCAATCTGGCCCTGCGCTGGTTGCGGGAGACCTTCCCACGGAACTCCA  789

      219 hrSerProGlyAlaIleHisTyrCys.AlaSerThrSerGlnGluSerIr  235
          |:::|||||:::::::::::::::::    :::|||::::::      ||
      790 CTCAGCCGGCTGTACTGTACTTCGCAGATGACGACAACACGTACAGTCTG  839

      235 pThrAspSerGlnValAspSerSerCysSerGlyGlnProIleHisLeuS  252
          |:::::   :::      ::::::   :::|||   |||      ||||
      840 GAGCTCTTGAAGAGATGCGCAGCACAAGAAGGGTGTCCGTGTGGCCTGT  889

      252 rpGlnPheLeuLysGluLeuLeuLeuLysPro  262
          ||   ::::||      :::   ::::::|||
      890 GGCCTTTGTTGCCGCCTTCGGTATGAGGCCC  921

Seq_comm. Seq-id_gpipatent_Cnq'6A.1'MPL seq_US-08-948-705-2

Seq_documentation_block:
; Sequence 2, Application US/08948705A
; Patent No. 6043084
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Matthew J.
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Old, Lloyd J.
;  TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
```

```
;  TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND METHODS FOR DIAGNOSING AND
;  TITLE OF INVENTION: TREATING COLON CANCER
;  FILE REFERENCE: [illegible]
;  CURRENT APPLICATION NUMBER: US/08/948,705A
;  CURRENT FILING DATE: 1997-10-10
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2885
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
US-08-948,705-2


alignment_scores:
            Quality:   91.50              Length:    201
              Ratio:  0.973                Gaps:     10
 Percent Similarity:  46.766   Percent Identity: 24.876

alignment_block:
 US-09-126,945B-2 x US-08-948,705-2   ..

 Align seq 1/1  to: US-08-948,705-2  from: 1  to: 2885

       10 SerValSerProSerHisLeuLeuLeuPro.ProAspThrValSerArg  26
          |||:::|||||||:|||      |||||| |||:::::::::::   :
      526 TCACTCTCAACCAACTCACACTTCACTGCCTCCCCCAAGCTGTGGACACACT 575

       27 hrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgAspTrp  43
          :|||:   :::::::|||:::
      626 CAGGAGGCCGAGAGGCAGGC.............................. 645

       44 IrpSerProSerProProAlaThrProGluGlnGlyLeuSerAlaPheLy 59
                        |||   :::|||      |||       |||
      646 ..............CCTCCAGTCCCTGCGGCAGGGTGGCACGCTGACCG 680

       59 rLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaL 76
             :::      :::|||||||              ||||||
      681 GCAAGTTCATGAGCACATCCTCTAT............TCCTGGCTGCCTG 718

       76 ysAlaProGlyAlaSerSerArgGluGluProProGlu.........Glu 89
          :::   |||:::::::::::::|||:::|||||||||||:::
      719 CTGGGCGTGGCACTGGAGGGCGACGGGAGCCCCCACGGGCATGCCTCCCT 768

       90 ProGluGluCysProVal.IleAspSerGlnAlaProAlaGlySerLeuA 106
                |||   ||| ::::::::::            :::::::|||
      769 GCTGCAGCAATGTGCTGTTGCTGGAGCAGGCCCGGCAGCAGCAGCACCCTCA 818

      106 spLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnVal 122
            ||||||                          ||||::
      819 TTGCTGTGCCA.............................CTCCACGGGCAG 841

      123 GlnSerMetValValGlyGlu...ValLeuLysAspIleGluThrAlaGly 138
          :::   :::|||:::::|||||   |||      ::::::::::|||:::
      842 TCCCCACTAGTGACGGGTGAACGTGTGGCCACCAGCATGCGGACGGTAGG 891

      138 slysLeu..............LeuAsnIleThr...AlaAspProMetA 149
          ||||||               ||||::   |||   :::::::::::
      892 CAAGCTCCCGCGGCATCGGCCCTGAGCCGCACTCAGTCCTCACCGCTGC 941

      149 spTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGln 165
            |||||||:::   :::|||:::   :::::      :::|||||||
      942 CGCAGAGTCCCCAGGCCCTGCAGCAGCTGGTCATGCAACAACAGCACCAG 991

      166 TyrArgLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGl 182
             |||
      992 CAGTTCCTG.......................................... 1000

      182 uLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGly 198
                      |||:::|||   :::|||:::      ||||||
     1001 ..............GAGAAGCAGAAGCAGCAGCAGCTACAGCTGGGC 1033

seq_name: /[illegible]/[illegible]/us-09-126-945B-2.MB seq [illegible] US-08-715-568A-2
```

```
seq_documentation_block:
; Sequence 2, Application US/08715568A
; Patent No. 5856463
;  GENERAL INFORMATION:
;    APPLICANT:  Prydz, Hans Peter Blankenborg
;    APPLICANT:  Brede, Gaute
;    TITLE OF INVENTION:  P2X8 I Kb_ypes and Uses in Disease
;    TITLE OF INVENTION:  Treatment
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lerner, David, Littenberg, Krumholz & Mentlik
;      STREET:  600 South Avenue West
;      CITY:  Westfield
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07090-1497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/715,568A
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Foley, Shawn P.
;      REFERENCE/DOCKET NUMBER:  P-P2X 3.0-002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908 654-5000
;      TELEFAX:  908-654-7866
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3471 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-715-568A-2


alignment_scores:
            Quality:   91.50              Length:    845
              Ratio:  0.610                Gaps:     23
 Percent Similarity:  43.478   Percent Identity: 28.768

alignment_block:
 US-09-126-945B-2 x US-08-715-568A-2   ..

 Align seq 1/1  to: US-08-715-568A-2  from: 1  to: 3471

        2 GlySerAlaSerProGlyLeuSerSerValSerProSer..... 14
          |||:::::::::||   |||   |||   :::||||||:::
     1770 GGGAACTGCAGTGCTGGGTGGAGTGTCCTGTCGCCTCAGGACCCTTTGGG 1819

       15 ..............................HisLeuLeuLeuProProA 21
                                       :::   ||||||||||||
     1820 ACAGTTACTTCTGGGACCCCCTTTCCTCCCACAGAGCCCTTCTCCCTCCTT 1869

       21 spThrValSerArgThr......GlyLeuGluLysAlaAlaAlaGly... 34
          |||   |||::::::::      :::|||:::   :::::::::|||
     1870 TCACACATTCCCATGCATCCTGATCCTTAACATTATGCTCCAGTGGGAGA 1919

       35 .........AlaValGlyLeuGluArgArgAspTrpSerProSerPro.. 47
                  |||   :::|||          |||||||:::   ||||||
     1920 CCCTGGTAGGCA'AAAGCTG...TGCCTTGACTGGACCCGTAGCCCCTG 1966

       48 ...............ProAlaThrProGluGlnGlyLeuSerAlaPhel 59
                        |||       |||:::|||   ||||||||:|||
     1967 GCTAGGTCGAAACAGCCCTCCACCTCCCAGCCAAGATCTGTCTTCCTTCA 2016

       59 yrLeuSerTyrPheAspMetLeuTyrProGluAspSerSer..... 72
             :::         ::::::      |||:::   ||||::
```

```
 2017 TGGTGCCTCCAGGGAGCCTTCCTGGTCCCAGGACCTCTGGTGGAGGGCCA 2066

   74 .....TrpAlaAlaLy 76
                        |||||||||
 2067 TGGCGTGGACCTTCACCCTCTGGACTGTGTGGCCATGCTGGTCATCGGC 2116

   76 sAlaProGlyAlaSer.SerArgGluGluProProGluGluProGluGln 92
      :::|||||||:::||| |||   ::::::       ::::::|||
 2117 TTGCCCAGGCTCCAGCCTCTCCAGATTCTGAGGGGTCTCAGCCCACCGCC 2166

   93 CysProValIleAspSerGluAlaProAlaGlySerLeuAspLeuValPr 109
         ::::::      :::       |||   |||||||:::
 2167 CTTGGTGCCTTCTTTGTAGAGCCCACGGCTACCTCCCTCTCC..... 2208

  109 oGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetV 126
                |||:::   ||||||||:   ||||||
 2209 .....CCCTTCCATGTCCATTCCATTCCCCAGGGGCTCCTCTCC 2248

  126 alValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsn 142
      :::|||   |||:::|||         |||||||:::
 2249 AACTGGGG...GTGGTTAAAGGGACGCCCACTGCTGCT..... 2288

  143 IleThr.AlaAspProMetAspTrpSerPro..... 152
      ||| ::::::         |||:::|||
 2288 ...ACCTGGGGAATGGGGCACCTGGGGCCCAAGGCAGAGGGAAGGGGTC 2330

  153 .....SerAsnValGluLysTrpLeuLeuTrpThrGluHisGln 165
           :::::|||:::   |||:::|||          :::
 2331 CTCCCGATTAGGGTCGAGTGTCAGCCTGGGTTCTA.....TCC 2368

  166 TyrArgLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGl 182
      :::    |||:::       |||:::
 2369 TTTGGTGCAGCCCCATTGCCTTTTCCCTTCAGG..... 2401

  182 uLeuCys.AlaMetSerGluGluGlnPheArgGlnArgSerProLeuGly 198
       |||||| |||   |||:::         |||:::|||
 2402 .CTCTGTTGCTCCCTCCTCTGCAGCTGCACGAAGGCGC..... 2438

  199 GlyAspValLeuHisAlaHisLeu...AspIleTrpLysSerAlaAlaIr 214
                      ||||||   :::   |||       ||||||||
 2439 ..................CATCTGGTGTCTGCATGGGTTGTTGGCAGCCTG 2478

  214 pMetLys.GluArgIleSerProGlyAlaIleHisTyrCysAlaSer... 229
      |   ::: ::::::   :::|||       :::|||   |||   |||
 2471 G...GAGTGATCACTGCACGCCC...ATCCTCCACACCTGCCCATCGTGC 2514

  230 .........................ThrSerGluGluSerTrpThrAs 247
                               :::|||   :::::::|||
 2515 ACACCCACCCATGGTGCACACCTGTAGTCCTCCATGAGGACATGGGAAGG 2564

  247 pSerGluValAspSerSerCysSerGlyGlnProIleHisLeu..... 261
         :::       :::   :::::::::      ||||||
 2565 TAGGAGTTGCCGCCCTGGGGCAGGGTCCCGGGCTGCTCACCTCTCCCCTT 2614

  262 .........................TrpGln......PheLeuLysGlu 277
                              ||:::       :::|||   :::
 2615 CTGCTGAGCTTCTGCGCACCCCTCCCTGGAACTTAGCCATACTGTGTGAC 2664

  278 LeuLeuLeuLysProHisSerTyrGly 296
      |||   ||||||||||       |||
 2665 CTGCCTCTGAAACCAGGGTGCCAGGGG 2691

Seq name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-368-590-1

Seq documentation block:
; Sequence 1, Application US/09/368,590
; Patent No. 6187563
; GENERAL INFORMATION:
;  APPLICANT: Solimena, Michele
;  TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR
;  TITLE OF INVENTION: AUTOANTIGENS OF AUTOIMMUNE DISEASES
;  FILE REFERENCE: 101918-200 (OCR 941)
;  CURRENT APPLICATION NUMBER: US/09/368,590
;  CURRENT FILING DATE: 1999-08-04
;  EARLIER APPLICATION NUMBER: 60/095,657
;  EARLIER FILING DATE: 1998-08-07
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 7812
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(6879)
;   NAME/KEY: unsure
;   LOCATION: (100)...(102)
;   NAME/KEY: unsure
;   LOCATION: (1021)...(1024)
;   NAME/KEY: unsure
;   LOCATION: (2266)...(2268)
US-09-368-590-1


alignment_scores:
            Quality:   91.00              Length:     328
              Ratio:   0.659               Gaps:      18
 Percent Similarity:   42.073   Percent Identity: 23.476

alignment block:
 US-09-126-945B-2 x US-09-368-590-1/rev   ..

 Align seq 1/1  to reverse of: US-09-368-590-1  from: 1  to: 7812

        8 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
          :::|||:::|||||               |||:::   |||   |||||
     3895 GCAGCTGCTCCAGGC..............CCCGGCGCACTGCCTCTGCC 3861

       19 oProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaV 36
          ||||      ::::::      :::       :::|||   |||
     3860 PCCGGGCTGGGCAGCGACGCCACGCGCCCGCGCGCTCCAGCACCTCCTC 3811

       36 alGlyLeuGluArgArgAspTrpSerProSer..... 46
          |||       :::         |||||||:::
     3810 CAGGCGCGGCCCACATGCGCCTGGATCTCCCGCCGCAGGCCCTGGTTCTTTT 3761

       47 ..............ProProAlaThrProGluGluGlyLeuSerAlaPh 58
                        ||||||||||:::        :::|||||||||:::
     3760 TGATGTGCTGCTGGACCGCCTGCAAACCGTTGCCTCGCTCTGTCTGCATG 3711

       58 eTyrLeuSerTyrPheAspMetLeuTyrProGluAsp..... 70
          :   ::::::                 ||||||
     3710 GCCAGTGGCA.............GCCGCTCCTGAACCCATGCCAGCTCCTC 3673

       71 ..........................SerSerTrp..... 78
                                   :::   |||
     3672 GTCCAGGTCGTGCGCCACCTGGTGCAACTCCTTGGAAGCCAGCAGCAAGC 3623

       74 ....................AlaAlaLysAlaProGlyAlaSerSerAr 88
                              |||   ||||||||||||
     3622 GGCGGCGCTCCTGCAACGCTCGAGCAGGCGCACCAGGCGCTCGCCCACC 3573

       88 qGluGluProProGluGluProGluGlnCysProValIleAspSerGlnA 100
          ::::::      :::   |||      |||          :::  |
     3572 GCGTTCTGCCGCTCACCCACGCTCCTTGCTCGCCGGCTCCAGCGGCAG 3523

      100 IaProAlaGlySer.LeuAspLeu.....ValPro..... 109
          |||||   :::::: |||:::|||           ||||||
     3522 CGCCGCCGTCTGCGCTGCAGCTCTCCCACCTCCGCGTTACCACTCCTCCA 3473

      110 .....GlyGlyLeuThrLeuGluGluHisSerLeuGlu.....G 121
                |||||   |||   |||   :::::::::      |
     3472 CCTGCGACCTCCATGGACTGCAGCTTCTTGAGCTGACTGTTGACAGTGGCC 3423

      121 InValGlnSerMetVal.....ValGlyGlu 129
          |||||:::|||   |||           |||   |||
```

```
     4422 AGGTCTCCTCCAGGGTCCACGTCTTGCAGCTGGCTCTCCATGTGAAGGAG 4373

      130 ValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIlethrAlaAs 146
          :::|||   :::::::::   |||         |||         |||
     4372 CTTCTTGTCCAGCTCAGCAAAGCTCTGCACCAGCTGGTCTGCTTGCTGG 4323

      146 pPro.............MetAspIrpSer.............ProSerAsnV 155
           ||             :::   ||||::               |||:::
     4322 GCCTCAAAGAGCTGCCATGCCTTGGGCCTGGTTGGTTGCTCTCCAGCTCCT 4273

      155 alGlnLysIrpLeuLeuIrpIhrGlu.HisGlnTyrArgLeuProProMe 171
          :::||   |||  :::      :::   :::  |||  |||||||:
     4272 CCCAGCACTGCGGATCTGGCCCAGCTTCTTCCGCACGGAGGCCGCCAGT 4223

      171 tGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerG 188
          :  :::     :::::||||||   :::          :::||||:
     4222 TGGGGCTTCTCCTGCATCAGTTGCTGGGCCCTCCCGCTCGATCTTCTCCA 4173

      188 IerGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuHisAla 204
          ::::::::::         |||          :::          ::::::
     4172 GCCACTCCTTATTCTCAGCCAGCTCGGCCATGAATGCCTGGTGCCGCAGT 4123

      205 HisLeu.....AspIleIrpLysSerAla..................... 212
          ||||||      ::::::|||   ||||::
     4122 CATCTCTTATGCAGCTTGTGGTTCTCCTCCCGCGTGCCATCCCGCGCCAT 4073

      213 ..........AlaIrpMetLysGluArgThrSerProGlyAlaIleHisT 229
                  :::|||:::::::      :::|||     :::   :::
     4072 CAGCATCTTCTCATGGATCCAGCCATCCAGCTCGTGGTAGTCTCGGAGGA 4023

      229 yrCysAlaSerThrSerGluGluSerIrpThrAspSerGlnValAspSer 242
            |||:::|||::          ||||||
     4022 AGTGCTGCAGCTCA.........AGTTGT..................... 3988

      243 SerCysSerGlyGlnProIleHisLeuIrp 252
          |||||||||         ||||||   |||
     3002 TCATGTAGC...TTTTGCATCCATTGCTGG 2976

seq_name: /cgn2_6/ftp-data/1/1/1_FA_COMP_seq_US-08-060-925A-2.fsa 2.

seq_documentation_block:
; Sequence 12, Application US/08060925A
; Patent No. 5439824
;  GENERAL INFORMATION:
;     APPLICANT:  Brantley, Mark
;     APPLICANT:  Laubach, Victor
;     TITLE OF INVENTION:  INCREASED EXPRESSION OF ALPHA 1
;     TITLE OF INVENTION:  ANTITRYPSIN IN EXPRESSION VECTORS THROUGH THE INCLUSION OF
;     TITLE OF INVENTION:  INTRON II
;     NUMBER OF SEQUENCES:  12
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  KNOBBE, MARTENS, OLSON AND BEAR
;       STREET:  620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
;       CITY:  NEWPORT BEACH
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  92660
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/060,925A
;       FILING DATE:  06-MAY-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fuller, Michael L.
;       REGISTRATION NUMBER:  36,516
;       REFERENCE/DOCKET NUMBER:  NIH040.001A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-235-8550
;       TELEFAX:  619-235-0176
;
;  INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10627 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-060-925A-12


alignment_scores:
            Quality:   91.00              Length:    264
              Ratio:  0.910                Gaps:     17
 Percent Similarity:  37.879   Percent Identity: 23.864

alignment_block:
 US-09-126-945B-2 x US-08-060-925A-12/rev   ..

 Align seq 1/1  to reverse of: US-08-060-925A-12  from: 1  to: 10627

       44 SerProSerProProAlaThr.............................. 50
          ||||||   ||||||:::|||
     8498 TCCCCGGTCCCCCCAGGACATTCTACCTTGAACTCCACACTCCACTGC 8449

       51 .....................................ProGluGlnG 54
                                              |||||||:::
     8448 TGCCTTTGCCAGGAAGCCCATCTGTTCCTTTTGGTTCTGCCAGAACGTG 8399

       54 lyLeuSerAlaPheTyrLeuSerTyr...PheAspMetLeuTyrProGlu 69
          :::     :::   |||   :::   :::|||:::          :::
     8398 TGGTGGTGCTGCTGTCCTGCCTGGGCACTGGATATTGGCAACGGACAG 8349

       70 AspSerSerIrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluPr 86
                  |||:::::::|||::|||||      :::|||      ||
     8348 TGTCCACACTGGACTGGGAAGTCCCCAGGG......ACGAGACCTTTACC 8305

       87 rrSerGluGlnProProGlnLysValIleAspSerGlnAlaProAlaG 103
          ||||   :::|||      ||||||                |||   |
     8304 TCCT...CACCCTGGGTACTGTCCT.................CCTCATG 8276

      103 lySerLeuAspLeuValPro.........GlyGlyLeuThrLeuGluGlu 116
          ||:::          :::|||        |||||||||   ||||||
     8275 GACCATGGATGTCGTCTCCTCAACTCACTGCTGGCCTCATTCTCGAA... 8229

      117 HisSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLysAs 133

     8229 ..................................................... 8229

      133 pIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspI 150
                    :::|||:::::::::::   :::::||||||      |
     8228 .............GCCAACTTTATACACACTACCACTGACCCACCCATCC 8192

      150 rp...SerProSerAsnValGlnLysIrpLeuLeuIrpThrGluHisGln 165
          ||   ||||||||||:::::              ||||||         :::
     8191 GGGGTTCACCCTCCTCAGCC.............CTCTGGCCAGTCCTGATG 8154

      166 TyrArgLeuProProMetGlyLys........................... 173
                ||||   ||||:|||
     8153 GGCCTCAGTCCCAACATGGCTAAGAGGTGTGGGCAGCTTCTTGGTCACCC 8104

      174 ..........................AlaPheGlnGluLeuAlaGlyL 181
                                   ||||||||||||   |||
     8103 TCAGGTTGGGGAATCACCTTCTGTCTTCATTTCCAGGAACTT...GGTG 8057

      181 ysGluLeuCysAla...................................... 185
          ::::::   :::
     8056 ATGATATCGTGGTGAGTTCATTTTCCAGGTGCTGTAGTTTCCCCTCATC 8007

      186 MetSerGluGluGln.....................PheArgGlnArgSe 195
          :::||||||:::                       |||:::|||:::
     8006 AGGCAGGAAGAAGATGGCGTGGCATTGCCCAGGTATTCCATCAGCAGCA 7957
```

```
      195 rProLeuGlyGly.............AspValLeuHisAla..... 204
          |||   |||                 ||||||   ||||||
     7956 CCCAGCTGGACAGCTTCTTACAGTGCTGGATGTTAAACATGCCTAAACGC 7907

      205 .........................HisLeu.AspIleTrpLysSerAl 212
                                   ||||||  :::   ||||||||||:
     7906 TTCATCATAGGCACCTTCACGGTGGTCACCTGGTCCACGTGGAAGTCCTC 7857

      212 aAlaTrpMetLysGluArgThrSerProGlyAlaIleHis 225
          ::::   :::   :::   ||||||   |||       |||
     7856 TTCCTCGGTG...TCCTGACTTCAAAGGTCTCTCCCAT 7820

Seq number, 2 to: 2 Protein 2 rna SA DMb seq      : 1-1 score: 0

Seq documentation block:
; Sequence 47, Application US/08131465B
; Patent No. 5527690
;  GENERAL INFORMATION:
;     APPLICANT:  Brown, Michael S.
;     APPLICANT:  Briggs, Michael R.
;     APPLICANT:  Wang, Xiaodong
;     APPLICANT:  Goldstein, Joseph L.
;     TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;     TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;     TITLE OF INVENTION:  PROTEINS
;     NUMBER OF SEQUENCES:  64
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  U.S.A.
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/131,465B
;       FILING DATE:  01 OCT 1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Parker, David L.
;       REGISTRATION NUMBER:  32,165
;       REFERENCE/DOCKET NUMBER:  UTSD.972/PAR
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418 3000
;       TELEFAX:  (512) 474 7577
;   INFORMATION FOR SEQ ID NO:  47:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4154 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "DNA"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  167..3607
US 08 131 465B 47


alignment scores:
            Quality:   90.50              Length:     810
              Ratio:  0.748                Gaps:      19
 Percent Similarity:  89.032   Percent Identity: 24.526

alignment block:
 US 09-126-945B 2 x US 08 131 465B 47  ..

 Align seq 1/1  to: US 08 131 465B 47  from: 1  to: 4154

      4 AlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProPr 20
        |||||||||      |||      |||||||::   |||      |||||
    850 GCCAGCCCCCATACCAGC......TCCCCAGGCAGCTTGTCTCCACCTCC 893

     20 oAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValG 37
        |   |||:::|||   :::::::||||||          ::::||     :
    894 TGCCACATTGAGC...TCCTCTCTTGAAGCCTTCCCTGAGCGGGCCGCAGG 940

     37 IyLeuGluArgArgAspTrpSerProSerProProAlaThrProGluGln 53
        ::                 :::||||||||||   :::   ||:::
    941 CA.................GCGCCCTCACCCCTGTCCCCTCCCCAGCCT 972

     54 ...yIeuSerAlaAsn...ProSerTyrPheAspMetLeuTyrProGluAs 70
        :::   :::   :::                  :::||||||
    973 GCACCCACTCCATTG.................AAGATGTACCCG..... 999

     70 pSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGlu.... 85
          |||      |||   :::||||||   :::   :::||||||
   1000 ....TCCATGCCGCTTTCTCCCCTGGGCCTGGTATCAAGGAAGAGTCAG 1045

     85 ..................................................... 85

   1046 TGCCACTGAGCATCCTGCAGACCCCCACCCCACAGCCCCTGCCAGGGCCC 1095

     86 .................ProProGluGluProGluGln.........Cy 98
                        |||         |||   |||
   1096 CTCCTGCCACAGAGCTTCCCAGCCCCAGCCCCACCGCAGTTCAGCTCCAC 1145

     98 sProValIleAspSerGlnAlaProAlaGly......SerLeuAspLeuV 108
        ||||||||:          :::|||   |||       |||
   1146 CCCTGTGTTAGGCTACCCCAGCCCTCCGGGAGGCTTCTCTACAGGAAGCC 1195

    108 alProGly.....................GlyLeuThrLeuGluGlu... 116
        ||||||                      ||||||   |||   :::
   1196 CTCCGGGAACACCCAGCAGCCGCTGCCTGGCCTGCCACTGGCTTCCCCG 1245

    117 ........................HisSerLeuGluGlnValGlnSerMe 125
                                |||:::      ||||||||||||||:
   1246 CCAGAGGTCCCGCCCGTCTCCTGCACACC.....CAGGTCCAGACTGT 1289

    125 tValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuA 142
        ::||      :::                              ||||||:
   1290 GGTCCCCCAGCAG.............................CTACTGA 1329

    142 snIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrp 158
        ::::::|||||   |||         ||||||   |||
   1330 CAGTCACAGCTGCCCCACGGC....ACCCCCTGTAAC.............. 1373

    159 LeuLeuTrpThrGluHisGlnTyrArgLeuProProMet......GlyLy 173
                 :::|||              |||       |||
   1374 ..........GACCACTGTGACCTCGCAGATCCAGCAGGTCCCGGTCC 1417

    173 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluG 190
        |||      |||          :::|||:::|||   ::::::::::
   1418 TGCTGCAGCCCCACTTCATCAAGGCAGACTCGCTGCTTCTGACAGCCATG 1467

    190 InPheArgGluArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu 206
        ::   |||:::            ||||||
   1468 AAGACAGACGGAGCCACTGTGAAGGCGGC..................... 1507

    207 AspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAl 223
                                :::      |||||||||
   1508 ..........................AGGTCTCAGTCCCCTGGTCT 1547

    223 alleHisTyrCysAlaSerThrSerGluGluSerTrpThrAsp....... 237
        |||:::||||||          ::::::::::::::|||
   1548 CTGGCACCACTGTGCA.......GACAGAGCCTTGCCGACCCTGTG 1591

    238 ........SerGluValAspSerSerCysSerGly.............. 246
                ::::::::   :::|||   ::|||
   1592 AGTGGCGCAACCATCTTGGCAACAGTCCCACTGGTCGTAGATGCGGAGAA 1641
```

```
     247 .........GlnProIleHisLeuTrpGln 258
                 ||||||   :::   ||||||
    1072 GCTGCCTATCAACCGGCTCGCAGCTGGCAG 1101

Seq_name: /tmp_1/pto/ptodir/1/fra_58_COMB.seq:US-08-668-123-37

Seq_documentation_block:
; Sequence 37, Application US/08668123
; Patent No. 5891631
;  GENERAL INFORMATION:
;     APPLICANT:  Brown, Michael S.
;     APPLICANT:  Briggs, Michael R.
;     APPLICANT:  Wang, Xiaodong
;     APPLICANT:  Goldstein, Joseph L.
;     TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;     TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;     TITLE OF INVENTION:  PROTEINS
;     NUMBER OF SEQUENCES:  64
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  U.S.A.
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/668,123
;       FILING DATE:  14-JUN-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/131,395
;       FILING DATE:  01-OCT-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Parker, David L.
;       REGISTRATION NUMBER:  32,165
;       REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (512) 474-7577
;   INFORMATION FOR SEQ ID NO:  37:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4154 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "DNA"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  167..3607
US-08-668-123-37


Alignment_scores:
            Quality:   90.50              Length:    310
              Ratio:   0.748                Gaps:     19
 Percent Similarity:   39.032   Percent Identity: 24.516

Alignment_block:
 US-09-126-945B-2 x US-08-668-123-37   ..

 Align seq 1/1  to: US-08-668-123-37  from: 1  to: 4154

        4 AlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProPr 26
          |||||||||      |||      |||||||::   |||      |||||
      850 GCCAGCCCGGATACCAGC......TCCCCAGGCAGCTTGCTCCACCTCC 893

       20 GAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValG 67
```

```
      |   |||:::|||   ::::::||||||         :::|||      :
    394 TGCCACATTGAGC...TCCTCTCTTGAAGCCTTCCTGAGCGGGCCGCAGG 440

     37 lyLeuGluArgArgAspTrpSerProSerProProAlaThrProGluGln 54
        ::          :::|||||||||   :::   ||:::
    441 CA.................GCGCCTCACCCCTGTCCCCTCCCCAGCCT 472

     54 GlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAs 70
        :::   :::   :::                 :::|||||
    473 GCACCCACTCCATTG.................AAGATGTACCCG..... 499

     70 pSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGlu.... 85
          |||       |||   :::|||||   :::   :::|||||
    500 ....TCCATGCCCGCTTTCTCCCCTGGGCCTGGTATCAACGAAGAGTCAG 545

     85 ..................................................... 85

    546 TGCCACTGAGCATCCTGCAGACCCCCACCCCACAGCCCCTGCCAGGGCCC 595

     86 ...................ProProGluGluProGluGln.........Cy 99
                           |||       |||   |||
    596 CTCCTGCCACAGAGCTTCCCAGCCCCAGCCCACCGCAGTTCAGCTCCAC 645

     99 sProValIleAspSerGlnAlaProAlaGly......SerLeuAspLeuV 108
          |||||||::          :::|||   |||       |||
    646 CCCTGTGTTAGGCTACCCCAGCCCTCCGGGAGGCTTCTCTACAGGAAGCC 695

    108 alProGly.....................GlyLeuThrLeuGluGlu... 116
          ||||||                    ||||||   |||   :::
    696 CTCCCGGGAACACCCAGCAGCCGCTGCCTGGCCTGCCACTGGCTTCCCCG 745

    117 .......................HisSerLeuGluGlnValGlnSerMe 125
                               |||:::      ||||||||||||::
    746 CCAGGGGTCCCCGCCCGTCTCCTTGCACAGC.....CAGGTCCAGAGTGT 789

    125 tValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuA 142
        :|||    :::                 |||||||:
    790 GGTCCCCCAGCAG.............................CTACTGA 809

    142 snIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrp 158
        ::::||||||   |||         ||||||   |||
    810 CAGTCACAGCTGCCCCCACGGC....AGCCCCTGTAAC............ 848

    159 LeuLeuIrpThrGluHisGlnTyrArgLeuProProMet......GlyLy 173
                  :::|||                |||         |||
    844 ...........GACCACTGGTGACCTCGCAGATCCAGCAGGTCCCGGTCC 891

    173 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluG 190
         |||         |||         :::|||:::|||   ::::::::::
    892 TGCTGCAGCCCCACTTCATCAAGGCAGACTCGCTGCTTCTGACAGCCATG 941

    190 InPheArgGluArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu 206
        ::   |||:::          ||||||
    942 AAGACAGACGGAGCCACTGTGAAGGCGGC..................... 969

    207 AspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAl 223
                                     :::     |||||||||
    961 ..............................AGGTCTCAGTCCCCTGGTCT 980

    223 alIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAsp..... 237
          |||::::||||          ::::::::::::::|||
    981 CTGGCACCACTGTGCA.........CACAGGCCTTTGCCGACCCTGGTG 1021

    238 ........SerGluValAspSerSerCysSerGly.............. 249
                :::::::::   :::|||   :::|||
   1022 AGTGGCGGAACCATCTTGGCAACAGTCCCACTGGTCGTAGATGCGGAGAA 1071

    247 .........GlnProIleHisLeuTrpGln 258
                ||||||   :::   ||||||
   1072 GCTGCCTATCAACCGGCTCGCAGCTGGCAG 1101

Seq_name: /tmp_1/pto/ptodir/1/fra_58A_COMB.seq:US-08-149-103-2
```

```
seq documentation block:
; Sequence 2, Application US/08149104
; Patent No. 5750367
;  GENERAL INFORMATION:
;     APPLICANT:  Lawrence C. B. Chan
;     TITLE OF INVENTION:  HUMAN AND MOUSE VERY LOW
;     TITLE OF INVENTION:  DENSITY LIPOPROTEIN RECEPTORS
;     TITLE OF INVENTION:  AND METHODS FOR USE OF SUCH
;     TITLE OF INVENTION:  RECEPTORS
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LYON & LYON
;       STREET:  611 West Sixth Street
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90017
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  4.5" Diskette, 1.44 Mb storage
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  IBM MS DOS (Version 5.0)
;       SOFTWARE:  WordPerfect (Version 5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/149,104
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;     PRIOR APPLICATION DATA:  including application
;     PRIOR APPLICATION DATA:  described below:                    none
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warburg, Richard J.
;       REGISTRATION NUMBER:  32,327
;       REFERENCE/DOCKET NUMBER:  204/052
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489 1600
;       TELEFAX:  (213) 955 0440
;       TELEX:  67 3510
;  INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4116 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US 08 149 104 2


alignment scores:
            Quality:    90.00              Length:     269
              Ratio:  0.796                Gaps:        11
 Percent Similarity:  42.007   Percent Identity: 20.818

alignment block:
 US 09 126-945B 2 x US 08 149 104 2/rev   ..

 Align seq 1/1  to reverse of US 08 149 104 2  from  1  to    4116

        3 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeu.. 18
          |||:::  |||:::|||:::|||  ||||||||||||      :::
      828 TCAATCGCTGCCCTTCTTGCAGTCAGCCTCTCCTCACACCCCCATTTTTT 779

       19 ..............ProProAspThrValSerArgThrGlyLeuGluL 50
                        |||     :::|||  :::|||
      778 GTGAATGCACTCGCCAGAGCCACACTGGATCTCACTGGTAGCACATTTGC 729

       51 ysAlaAlaAlaGly.........AlaValGlyLeuGluArgArgAspTrp 83
            :::   :::|||          :::   :::      ||||||
      728 TATGTATCACAGCCTGACCGCCACACTGCCTCAAGACACTCCTCTGATTGG 679

       84 .........SerProSerProProAlaThrProGluGlnGlyLeuSerAl 117
                  |||  |||        |||      :::|||:::::
      678 TCTGAACAGTCTGCTCATCATCACACCCAGCTGAGGGCAATGCAGGA 629

       57 aPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpA 74
                |||  :::                 :::|||||||
      628 AGAGGTGCTGCACTGG.....................AACTCGTGGG 603

       74 laAlaLysAlaProGlyAlaSerSerArgGluGluProProGluGluPro 90
          ||   ::::::   |||||||:::|||   ::::::   |||:::::::
      602 CTCCGCACGCTGGTGAGCACACGTCCAGCTCATCACTGCCATCGTCACAG 553

       91 GluGlnCysProVal.............IleAspSerGlnAlaProAlaGl 103
          ::::::   ||||::             ::::::::|||   |||
      552 TCATCCTGGCCATTGCACACAAAGTTCTGGACACGCAGCGGCCA..... 508

      103 ySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuG 120
                                                      ||||
      507 ..................................................CTGG 504

      120 luGlnValGlnSerMetValValGlyGluValLeuLysAspIleGluThr 136
          |||||||:::|||   ::::::     :::|||  ::::::::::::
      503 AGCAAGTGAACTCATCTGCACTACATGTTATGTTGCCACAGTTTTCTTCA 454

      137 AlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSe 153
          :::     |||                 :::::::::|||||||
      453 TCTTCTCCATTG.................TCACAATCATTTTCACCATC 422

      153 rAsnValGluLysTrpLeuLeuTrpThrGluHisGluTyrArgLeuProP 170
          |::::::|||::      :::          |||      :::      }
      421 GCATCTCAGGAGACGGGATA.........CACTGAGTAGAACGGGCGC 381

      170 roMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMet 186
          ||:::     :::||:::::::
      380 CACAGCTGATTTCATTTATGCGG.......................... 358

      187 SerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuRi 203

      358 .................................................. 358

      203 sAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrS 220
              |||:::   ||||||
      357 ....CATGTTCTCATATGG.............................. 343

      220 erProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThr 236
                        :::|||:::::::   |||:::|||   ::::::
      342 ................CACTGTTCAGGGCTTTCATCAGAACCGTTTTCG 310

      237 AspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGl 253
          :::|||       |||   |||                 |||   |||
      309 CAATCAGGATCCCCGTCACACTGCCATCTGTTAGGAACACAC...TGGCC 263

      253 nPheLeu 255
            ||||||
      262 GTTTTG 256

Seq_name: 5940.2 /pf/data/Cori5A_5Mo/seq.DC v8 451 888 2

seq_documentation_block:
; Sequence 2, Application 08/08451884
; Patent No. 5798209
;  GENERAL INFORMATION:
;     APPLICANT:  Lawrence C.B. Chan
;     TITLE OF INVENTION:  HUMAN AND MOUSE VERY LOW DENSITY
;     TITLE OF INVENTION:  LIPOPROTEIN RECEPTORS AND METHODS FOR
;     TITLE OF INVENTION:  USE OF SUCH RECEPTORS
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LYON & LYON
;       STREET:  633 West Fifth Street, Suite 4700
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90071-2066
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  4.5" Diskette, 1.44 Mb storage
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  IBM MS-DOS (Version 6.22)
;       SOFTWARE:  WordPerfect (Version 5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/451,883
;       FILING DATE:  May 26, 1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;     PRIOR APPLICATION DATA:  including application
;     PRIOR APPLICATION DATA:  described below:
;       APPLICATION NUMBER:  08/149,103
;       FILING DATE:  No. 5798209ember 8, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Knight, Matthew W.
;       REGISTRATION NUMBER:  36,846
;       REFERENCE/DOCKET NUMBER:  212/268
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489 1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3116 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-451-883-2


alignment_scores:
            Quality:   90.00              Length:    269
              Ratio:   0.796               Gaps:     11
 Percent Similarity:   42.007   Percent Identity: 20.818

alignment_block
 US-09-126-945B-2 x US-08-451-883-2/rev   ..

 Align seq 1/1  to reverse of: US-08-451-883-2  from: 1  to: 3116

        3 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeu.. 18
          ||:::  |||:::|||:::|||  |||||||||||      :::
      828 TCATCGCTGCCGTCCTTGCAGTCAGGGTCTCCGTCACACGGCCATTTTT 779

       19 ..............ProProAspThrValSerArgThrGlyLeuGlut 38
                        |||     :::|||  ::|||
      778 GTGAATGCACTCGCCAGAGCCACACTGGATCTCACTGGTAGGACATTTGG 729

       39 ysAlaAlaAlaGly.....AlaValGlyLeuGluArgArgAspTrp 63
          :::  :::|||            :::  :::     ||||||
      728 TATGTATCACAGGCTGACGGCCACACTGCTCAAGAGACTCGTCTGATTGG 679

       64 .....SerProSerProProAlaThrProGluGlnGlyLeuSerAl 87
                |||  |||       |||     :::|||:::::
      678 TCTGAACAGTCTGCGTCATCATCACACCCAGCTGACGGGAATGCAGGA 629

       88 aPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpA 114
             |||  :::                    :::|||||||
      628 AGAGGTGCTGCACTGG.....................AACTCGTGCG 603

      115 laAlaLysAlaProGlyAlaSerSerArgGlyProProPheGlyPr 144
          ||  ::::::  |||||:::|||  ::::::  |||::::::
      602 CTCCGCAGGTTGGTGGAGCACAGTCCAGCTCATCACTGCCATCGTCACAG 553

      145 GluGlnCysProVal.....IleAspSerGlnAlaProAlaGl 168
          ::::::  |||:::          :::::::::|||  |||
      552 TCATCCTGGCCATTGCACACAAAGTTTCTGGAGACGCAGCGGCCA..... 509

      169 ySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuG 193
                                                 ||||
      508 .....CTGG 504

      193 luGlnValGlnSerMetValValGlyGluValLeuLysAspIleGluThr 138
          ||||||||:::|||  ::::::      :::|||  ::::::::::::
      503 AGCAAGTGAACTCATCTGCACTACATGTTATGTTGCCACAGTTTTCTTCA 454

      137 AlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSe 153
          :::      |||                    :::::::::|||||||
      453 TCTTCTCCATTG.................TCACAATCATTTTCACCATC 422

      152 rAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProP 170
          |:::::|||:::      :::       |||      :::      :
      421 GCATCTCCAGGAGACAGGGATA.........CACTGAGTAGAACGGGCGC 381

      170 roMetGlyLysAlaPheGluGluLeuAlaGlyLysGluLeuCysAlaMet 186
          |||:::     :::|||:::::
      380 CACAGCTGATTTCATTTATGCGG.......................... 358

      187 SerGluGluGlnPheArgGluArgSerProLeuGlyGlyAspValLeuHi 203

      358 .................................................. 358

      203 sAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgIhrS 220
              |||:::   ||||||
      357 ....CATGTTCTCATATGG............................. 343

      220 erProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpIhr 236
                        :::|||:::::::   |||:::|||   ::::::
      342 ..............CACTGTTCAGGGCTTTCATCAGAACCGTTTTCG 310

      237 AspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGl 253
          :::|||       |||   |||                |||   |||
      309 CAATCAGGATCCCCGTCACACTGCCATCTGTTAGGAACACAC...TGGCC 263

      253 nPheLeu 255
          ||||||
      262 GTTTTG 256

seq_documentation_block:
; Sequence 47, Application US/08630916A
; Patent No. 6011137
;  GENERAL INFORMATION:
;    APPLICANT:  Pirozzi, Gregorio
;    APPLICANT:  Kay, Brian K.
;    APPLICANT:  Fowlkes, Dana M.
;    TITLE OF INVENTION:  IDENTIFICATION AND ISOLATION OF NOVEL
;    TITLE OF INVENTION:  POLYPEPTIDES HAVING WW DOMAINS AND METHODS OF USING SAME
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/630,916A
;      FILING DATE:  03-APR 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MISRock, S. LESLIE
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  1101-203
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3479 bases
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-630-916A-47


alignment scores:
            Quality:   90.00              Length:     279
              Ratio:   0.738                Gaps:      16
 Percent Similarity:   47.104   Percent Identity: 22.394

alignment block:
 US-09-126-945B-2 x US-08-630-916A-47   ..

 Align seq 1/1  to: US-08-630-916A-47  from: 1  to: 3476

       13 ProSerHisLeuLeuLeuProProAspThrValSerArgThrGlyLeuCl 29
          ||||||||||:::   :::          :::||||||||||      |||
     1682 CCTAGCCACGTGAAGATC.........AGCGTTTCCAGGCAGACGCTTAT 1722

       29 uLysAlaAlaAlaGlyAlaValGly......LeuGluArgArgAspTrpS 44
          :::||||||||||   ::::::    :::   ||||||   |||:
     1723 GACCTGCGCCGCCGGCTTTACATCATGCCTGGCGGAGGAGGGCCTGGA 1772

       44 crProSerProProAlaThrProGluGlnGlyLeuSerAla..... 57
          ::   :::      |||:::||||||:::||||:::||||
     1773 CTATGGGG......GCATCGCCAGAGACTGGTTTTTCCTCCTGTCTCAGG 1816

       58 ............PheTyrLeuSerTyrPheAspMetLeuTyrProGluAs 70
                     :::   :::   |||:::::|||      ||||||
     1817 AGGTGCTCAACCCTATGTATTGTTTATTTGAATATG......CCGCAAAG 1860

       70 pSerSerTrpAlaAlaLysAla......ProGlyAlaSerSerArgGluG 85
          :::      |||:::::::::      |||      |||:::|||
     1861 AACAATTACTGCCTGCAGATCAACCCCGCCTCCTCCATCAACCCGGACCA 1910

       85 luProPro.....................GluGluProGluGlnCysPro 94
             |||                    :::::|||   :::|||
     1911 CCTCACCTACTTTCGCTTTAIAGGCAGATTCATCGCCATGGCGCTGT... 1957

       95 ValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGl 111
          ::::::::::|||:::::      :::|||   :::|||      :::||
     1958 ACCATGGAAAGTTCATCGACACGGGCTTCACCCTCCCTTTCTACAAGCGG 2007

      111 yLeuThrLeuGluGluHisSerLeuGluGlnValGluSerMetValVal 128
          |  :::::::::::::::      :::      :::   :::
     2008 ATGCTCAATAAGAGACCAACCCTGAAAGACCTGGAGTCCATTG....... 2050

      128 lyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThr 144
             :::|||:::::                          :::   |||
     2051 .....ACCCTGAGTTCTACAACTCCATTGTCTGGATCAAACAGAACAACC 2095

      145 AlaAspProMetAspTrpSerProSerAsnValGluLysTrpLeuGlnTr 161
                :::   ||||||   ::::::   :::   |||      ||
     2096 TGGAAGAATGTGGCCTGGAGCTGTACTTCATCCAGGACATGGAGATACTG 2145

      161 pThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGluGluL 178
          |:::          |||   ||||||
     2146 GGCA.............AGGTCACCACCC.....................A 2162

      178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGluArg 194
             :::   ::::::   :::      |||   ::::::   |||   |||
     2163 CGAGCTGAAGGAGGGCGGCGACAGCATCCGGGTCACGGAGGAGAACAAGG 2212

      195 ..............................SerProLeuGlyGlyAs 200
                                        ||||||
     2213 AACAGTACATCATGCTGCTGACTGACTGGCGTTTCACCC.......... 2251

      200 pValLeuHisAlaHisLeuAspIleTrpLysSer.........AlaAlal 214
                        :::   |||||||||              :::|
     2252 ..................GAGGCGTGGAAGAGCAGACCAAAGCCTTCCT 2282

      214 rpMetLysGluArgThrSerPro.........GlyAlaIleHisTyrCys 227

          |||||   :::          |||          |||          |||
     2283 GGATGGCTTCAACGAGGTGGCCCCGCTGGAGTGGC.............TGC 2320

      228 AlaSerThrSerGluGluSerTrpThr 236
          |||:::   :::::::::|||||||:::
     2321 GCTACTTTGACGAGAAAGAGCTGGAGC 2347


seq_name _/ n=2_/pf dir=1_fil=1_"A_" MuSeq US-08-342-930-1

seq_documentation_block:
; Sequence 1, Application US/08342930
; Patent No. 5821084
;  GENERAL INFORMATION:
;     APPLICANT:  OLMSTED, ELIZABETH A.
;     APPLICANT:  MAURO, LAURA J.
;     APPLICANT:  DAVIS, ALAN R.
;     APPLICANT:  DIXON, JACK E.
;     TITLE OF INVENTION:  OSTEOBLAST TESTICULAR PROTEIN TYROSINE
;     TITLE OF INVENTION:  PHOSPHATASE
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORRISON & FOERSTER
;       STREET:  755 Page Mill Road
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94304-1018
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/342,930
;       FILING DATE:  21-NOV-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  KONSKI, ANTOINETTE F.
;       REGISTRATION NUMBER:  34,202
;       REFERENCE/DOCKET NUMBER:  20344-20975.00
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 813-5600
;       TELEFAX:  (415) 494-0792
;       TELEX:  706141
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5455 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  205..5337
US-08-342-930-1


alignment_scores:
            Quality:   90.00              Length:     814
              Ratio:   0.677                Gaps:      14
 Percent Similarity:   42.492   Percent Identity: 21.406

alignment_block:
 US-09-126-945B-2 x US-08-342-930-1   ..

 Align seq 1/1  to: US-08-342-930-1  from: 1  to: 5455

        9 SerSerValSerProSerHisLeuLeu...LeuProProAspThrValSe 24
          |||:::   ||||||||:::   |||:::   |||   ||||||||||:::||
     1741 AGCCAGAGGTCTCCTGGCGACCTTGTTGACTTGGGCCCTGACACTTTGAG 1790

       24 rArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgA 41
          |   |||      :::   :::   :::::::::      ::::::
     1791 CCTGACCTCTGAAATCTCTGGTAGCCGGCTCCTGCTACACCGTGTCAGCAT 1840
```

```
     41 rqAspTrpSer..................................... 44
           |||:::
   1841 GGGCCTGGGCCGGGAACCTCAGCTCTGACTCTCAGAAGATTCACAGCTGC 1890

     45 ......ProSerProProAlaThr....................... 59
              |||:::||||||:::::
   1891 ACCCGCCCCGCTCCTCCCACCAACCTGAGTCTGGGCTTTGCCCACCAGCC 1940

     51 ........................ProGluGlnGlyLeuSerAlaP 68
                                 |||       |||   :::|||
   1941 TGCGGCACTGAAGGCTTCCTGCTATCACCCACCGGGTGGCAGGGATGCCT 1990

     68 heTyrLeuSerTyrPhe........................AspMet 65
        |||:|||      :::                       :::
   1991 TTCACTTACGGCTTTACAGGCTGAGGCCTCTGACACTGGAAAGTGAGAAG 2040

     66 LeuTyrProGluAspSer.........SerTrpAlaAlaLysAlaProGl 79
        :::   |||:::::::::::      ||||||||||      :::   ||
   2041 CTCCTACCTCCGGAGGCCACAACTTCTCCTGGGCCACTGCTAAGCCTGCAA 2090

     79 y.........................................AlaS 81
        |                                          :::|
   2091 CTGTGAGTTCCAGGTACAGCTGTCTACCTTCTGGGGGTCTGAGAGAAGCA 2140

     81 erSerArqGluGluProProGluGluProGluGlnCysProVal..... 95
        |||||   :::                 |||   ::::::::::::
   2141 GCAGTGCCAAGGCCACAGGCTGGACACCCCCTTCAGCTCCTACACTGGTA 2190

     96 ...IleAspSerGlnAlaProAlaGly.........SerLeuAspLeuVa 108
           :::   ||||:::|||||||:::                 |||      ||
   2191 AACGTGACCAGCGATGCTCCTACCCAGCTCCAAGTATCCTGGGCCCACGT 2240

    108 IProGlyGlyLeuThrLeuGluGlnHisSerLeuGluGlnValGlnSerM 125
        ||||||||||   :::     :::   ::|||   |||   ::::::
   2241 TCCTGGGGGCCGGAGCCGCTACCAAGTGACCCTATACCAGGAGAGTACCC 2290

    125 etValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeu 141
          ::::::   :::::::::          |||   :::   :::::::||
   2291 GGACAGCCACCAGCATCATGGGCCCAAGGAAGATGGCACGAGCTTTTTG 2340

    142 AsnIleThr.....................................AlaAs 149
        ::::::|||                                    |||
   2341 GGTTTGACTCCTGGCACTAAGTACAAGGTGGAAGTCATCTCCTGGGCTGG 2390

    146 pProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrG 163
         |||:::     :::   :::||||||:::   |||
   2391 GCCCCTCTACACTGCAGCAGCCAACGTTTCTGCCTGG............. 2427

    163 luHisGlnTyrArqLeuProProMetGlyLysAlaPheGlnGluLeuAla 179
              |||   |||   |||                :::      |||
   2428 .....ACCTACCCACTCATACCCAATGAGCTGCTCGTGTCAATGCAGGCA 2472

    180 GlyLysGluLeuCysAlaMetSerGluGluGlnPheArqGlnArqSerPr 196
        |||:::   :::     ::::::      :::             ::::|
   2473 GGCAGTGCTGTGGTTAACCTGGCCTGGCCCAGT...........GGTCC 2519

    196 oLeuGlyGlyAspValLeuHisAlaHisLeuAsp..IleTrpLysSerAl 212
        |||||||      :::   |||||||:::|||:::     |||
   2511 CCTGGGGCAAGGGGCATGCCACGCCCAACTCTCAGATGCTGGACACCTCT 2560

    212 aAlaTrpMetLysGluArqThrSerProGlyAlaIleHisTyrCysAlaS 229
                      :::|||   |||   |||:::|||      :::|
   2561 CATGGGAGCAACCCCTGAAACTAGGCCAAGAGCTCTTCATGCTAAGGGAT 2610

    229 erThrSerGluGluSerTrpThrAsp...SerGluVal 240
        ||   ::::::             |||   ||||||||
   2611 CTCACACCAGGACATACCATCTCGATGTCAGTGAGGTG 2648

seq_name: /cgn2_6/ptodata/1/ima/58_COMB_seq/US-08-970-269A-32

seq_documentation_block:
; Sequence 32, Application US/08970269A
; Patent No. 5976803
;  GENERAL INFORMATION:
;    APPLICANT:  Katheryn Meek
;    TITLE OF INVENTION:  Genetic Test For Equine Severe
;    TITLE OF INVENTION:  Combined Immunodeficiency Disease
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dr. Benjamin A. Adler
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple
;      OPERATING SYSTEM:  MacinTosh
;      SOFTWARE:  Microsoft Word for Macintosh
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/970,269A
;      FILING DATE:  No. 5976803ember 14, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler Ph.D., Benjamin A.
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5860
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-777-2321
;      TELEFAX:  713 777 6908
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11873 bp
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  other nucleic acid
;    HYPOTHETICAL:  no
;    ANTI SENSE:  no
;    ORIGINAL SOURCE:
;    FEATURE:
US-08-970-269A-32


alignment_scores:
            Quality:   90.00              Length:     816
              Ratio:   0.657               Gaps:      12
 Percent Similarity:  43.354   Percent Identity: 18.048

alignment_block:
 US-09-126-945B-2 x US-08-970-269A-32    ..

 Align seq 1/1  to: US-08-970-269A-32  from: 1  to: 11873

      59 TyrLeuSerTyrPheAspMetLeuTyr..................... 67
         :::|||||   |||   :::|||||||
    8043 TTTCTTTCCACCTTTCATCTCCTGTATCCAGGAAATTAGTTGCCAACACG 8092

      68 ............ProGluAspSerSerTrpAlaAlaLysAlaProGly. 79
                     |||:::   |||         ::::::   ||||||
    8093 CAGACTTGCTGAGCCTCGACCCAGCTTCTGTCAGTGCCAGCTGCCTGGCC 8142

      80 .....AlaSerSerArqGluGluProProGluGluProGluGlnCysPro 94
           |||:::::   ::::::|||||:::             :::|||
    8143 AGTCTGCAGCAGCCTGTAGGCGTCCGCCTTCTGGAGGAGGCCTTGCTCCA 8192

      95 ValIleAspSerGlnAlaProAla.......................... 102
         ::::::   ::::::   |||||
    8193 CTGCTGCCTGAAGAGCCACCTGCCAAGCGAGTTCGAGGCAGACCCTGTCT 8242

     102 ..................................................... 102

    8243 CTAGCCTGATTTGTCAGATGGATGGAACTTGCTAAACTGTATAGATCAA 8292
```

```
  103 ..GlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSer 118
        ||||::   |||||::::::   |||   ::::::   |||      :::
 8293 TTGGAGAATATGACATCCTCCGTGGGATTTTTAATAGTGAGATAGGAACA 8342

  119 LeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspIleGl 135
        :::   :::||||::   :::::::::|||::   :::|||   ::
 8343 AAGCAAGTCACTCAGAATGCATTATTAGCAGAAGCAAGAAATGATTATTC 8392

  135 nThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerP 152
      :  |||  |||      |||  :::  :::  :::||||||
 8393 TGAAGCCGTTAAGCAGTATAATGAGGCTCTCAATAAACAAGACTGGGTAG 8442

  152 roSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeu 168
        ::::::
 8443 ATGGTGAG.................................................. 8460

  169 ProProMet.....GlyLysAlaPheGlnGluLeuAlaGlyLysGluLe 183
        ||||||          |||  |||  ||||||||||::   :::
 8461 ...CCTATGGAAGCTGAGAAGGATTTTTGGGAACTTGCATCCCTTGAC.. 8495

  184 nCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyA 200

 8496 .................................................... 8495

  200 spValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLys 216
            :::   :::|||      |||||||||   |||
 8496 .....TGTTATAACCAACTTGCTGAGTGGAAATCACTGGCA......... 8541

  217 GluArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGl 233
                              ||||||::::::::::|||   ::
 8542 ..........................TACTGTCTACAGTCAGTGTTGA 8554

  233 nSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleR 250
      :|||   :::      ::::::::::::      :::   :::|||:::
 8555 CAGTGCGAACCCTCCAGATTTAAATAAAATGTGGAAT...GAACCATTTT 8601

  250 isLeuTrpGlnPheLeuLys.....................GluLeuLeu 259
      ::         :::|||                      :::||||||
 8602 ATCAGGACACCTATCTACCTTACATGATCCGCAGCAAGCTGAAGCTACTT 8651

  260 LeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys..... 274
      ||||::   ::::::          ::::::   ::::::::::::
 8652 CTGCAAGGTGAGGGAGACCAGTCCCTGCTGACATTTATTGATGAAGCTGT 8701

  275 .............GluLysGlyIlePheLysIleGluAspSerAlaGlnV 287
                  :::|||   :::   :::::::::   |||   ::::
 8702 GAGCAAGGAGCTCCAGAAGGTCCTCGTAGAGCTTCATTACAGTCAGGAAT 8751

  287 alAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAsp 303
      :::::   ||||::   ::::::                      |||
 8752 TGAGTCTCCTTTATATCCTACAA.....................GA1 8777

  304 LysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArg 319
        ::::::||||::       :::|||||   ::::::   ||||||
 8778 GACGTCGACAGAGCC......AAATATTATATTGAAAATTGCATTCGG 8819

Seq name: /cm7/bpfiles/11nq59b_CoMb.seq:US-08-970-269A-41

seq_documentation_block:
; Sequence 41, Application US/08970269A
; Patent No. 5976803
;  GENERAL INFORMATION:
;     APPLICANT:  Katheryn Meek
;     TITLE OF INVENTION:  Genetic Test For Equine Severe
;     TITLE OF INVENTION:  Combined Immunodeficiency Disease
;     NUMBER OF SEQUENCES:  42
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Dr. Benjamin A. Adler
;       STREET:  8011 Candle Lane
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77071
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Apple
;       OPERATING SYSTEM:  Macintosh
;       SOFTWARE:  Microsoft Word for Macintosh
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/970,269A
;       FILING DATE:  No. 597680tember 14, 1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Adler Ph.D., Benjamin A.
;       REGISTRATION NUMBER:  35,423
;       REFERENCE/DOCKET NUMBER:  D5860
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  713-777-2321
;       TELEFAX:  713-777-6908
;   INFORMATION FOR SEQ ID NO:  41:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  11878 bp
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double stranded
;       TOPOLOGY:  linear
;     MOLECULE TYPE:
;       DESCRIPTION:  other nucleic acid
;     HYPOTHETICAL:  no
;     ANTI SENSE:  no
;     ORIGINAL SOURCE:
;     FEATURE:
US-08-970-269A-41


alignment_scores:
            Quality:   90.00              Length:    316
              Ratio:   0.657               Gaps:     12
 Percent Similarity:  43.354   Percent Identity: 18.038

alignment_block:
 US-09-126-945B-2 x US-08-970-269A-41    ..

 Align Seq 1/1  to: US-08-970-269A-41  from: 1  to: 11878

      59 TyrLeuSerTyrPheAspMetLeuTyr..................... 67
         :::||||||   |||   :::||||||
    8043 TTCTTCCACCTTTCATCTCCTGTATCCAGGAAATTAGTTGCCAACACG 8092

      68 .............ProGluAspSerSerTrpAlaAlaLysAlaProGly. 79
                      ||||::   |||        ::::::   ||||||
    8093 CAGACTTGCTGAGCCTCGACCCAGCTTCTGTCAGTGCCAGCTGCCTGGCC 8142

      80 .....AlaSerSerArgGluGluProProGluGluProGluGlnCysPro 94
              ||:::::::   ::::::||||||::          ::::||
    8143 AGTCTGCAGCAGCCTGTAGGCGTCCGCCTTCTGGAGGAGGCCTTGCTCCA 8192

      95 ValIleAspSerGluAlaProAla.......................... 102
         ::::::   ::::::   ||||||
    8193 CTGCTGCCTGAAGAGCCACCTGCCAAGCGAGTTCGAGGGAGACCCTGTCT 8242

     102 ..................................................... 102

    8243 CTACCCTGATTTTGTCAGATGGATGGAACTTGCTAAACTGTATAGATCAA 8292

     103 ..GlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSer 118
           ||::   |||::::::   |||   ::::::   |||      :::
    8293 TTGGAGAATATGACATCCTCCGTGGGATTTTTAATAGTGAGATAGGAACA 8342

     119 LeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspIleGl 135
           :::   :::||||::   :::::::::|||::   :::|||   ::
    8343 AAGCAAGTCACTCAGAATGCATTATTAGCAGAAGCAAGAAATGATTATTC 8392

     135 nThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerP 152
         :  |||  |||      |||  :::  :::  :::||||||
    8393 TGAAGCCGTTAAGCAGTATAATGAGGCTCTCAATAAACAAGACTGGGTAG 8442
```

```
      152 roSerAsnValGluLysIrpLeuLeuTrpThrGluHisGlnIyrArqLeu 168
            ::::::
     8444 ATGGTGAG.......................................... 8450

      169 ProProMet......GlyLysAlaPheGlnGluLeuAlaGlyLysGluIo 183
             |||:||      |||   |||   |||||||||||:   :::
     8451 ...CCIATGGAAGCTGAGAAGGATTTTTGGGAACTTGCATCCCTTGAC.. 8495

      184 qCysAlaMetSerGluGluGlnPheArqGlnArqSerProLeuGlyGlyA 200

     8496 .......................................................... 8499

      200 spValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaIrpMetLys 216
                  :::   :::|||      |||||||||   |||
     8496 .....IGTTATAACCAACTTGCTGAGTGGAAATCACTGGCA......... 8531

      217 GluArqIhrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGl 233
                                  |||||||:::::::|||   ::
     8532 ........................TACTGTTCTACAGTCAGTGTGGA 8554

      233 qSerIrpIhrAspSerGluValAspSerSerCysSerGlyGlnProIleH 250
          ::||   :::      ::::::::::      :::   :::|||:::
     8555 CAGTGTGAACCCTCCAGATTTAAATAAAATGTGGAAT...GAACCATTTI 8601

      250 isLeuIrpGlnPheLeuLys.....................GluLeuLeu 259
          ::        :::|||                        :::|||||
     8602 ATCAGGAGACCIAICIACCTIACATGATCCGCAGCAAGCTGAAGCIACTI 8651

      260 LeuLysProHisSerIyrGlyArqPheIleArqIrpLeuAsnLys..... 274
          ||:::  :::::  :::::   ::::::::::::
     8652 CTGCAAGGTGAGGGAGACCAGTCCCTGCTGACATTTATTGATGAAGCTGI 8701

      275 .............GluLysGlyIlePheLysIleGluAspSerAlaGlnV 287
                     ::|||   :::   :::::::::   |||   ::::
     8702 GAGCAAGGAGCTCCAGAAGGTCCTCGTAGAGCTTCATTACAGTCAGGAAI 8751

      287 qIAlaArqLeuIrpGlyIleArqLysAsnArqProAlaMetAsnIyrAsp 303
          :::::   ||:::   :::::                        |||
     8752 TGAGICTCCTTTATATCCTACAA.....................GAI 8777

      304 LysLeuSerArqSerIleArqGlnIyrIyrLysLysGlyIleIleArq 319
            ::::::|||:::      :::|||||   :::::   ||||||
     8778 GACGTCGACAGAGCC......AAATATTATATTGAAAATTGCATTCGG 8819

seq_name: /cgn2_6/ptodata/1/ima/5A_COMR seq:US-08-035-928-1

seq_documentation_block:
; Sequence 1, Application US/08035928
; Patent No. 5538844
;  GENERAL INFORMATION:
;    APPLICANT:  Duyao, Mabel P.
;    APPLICANT:  MacDonald, Marcy E.
;    APPLICANT:  Gusella, James F.
;    TITLE OF INVENTION:  A No. 5538844el Transport Protein Gene from
;    TITLE OF INVENTION:  the Huntington's Disease Region
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1225 Connecticut Avenue N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/035,928
;      FILING DATE:  19930323
;      CLASSIFICATION:  435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 466-0800
;      TELEFAX:  (202) 833-8716
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1788 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  121..1488
US-08-035-928-1


alignment_scores:
            Quality:   89.50              Length:    293
              Ratio:  0.694               Gaps:       17
 Percent Similarity:  44.027   Percent Identity: 24.573

alignment_block:
 US-09-126-945B-2 x US-08-035-928-1   ..

 Align seq 1/1  to: US-08-035-928-1  from: 1  to: 1788

        2 GlySerAlaSerPro..............GlyLeuSerSerValSerPr 18
          |||:::::::|||||              |||:::::::::::::   ||
      104 GGTCAGAGCAGCCACCATGGGATGGGGAGGGGGTGGAGGCTGCACCCCC 153

       18 oSerHisLeuLeuLeuProProAspIhrValSerArqIhrGlyLeuGluL 80
          |:::|||            |||:::|||   ||||||            :
      154 CGCCCAC............CCATCCACCAGCAGCCGC............C 179

       80 ysAlaAlaAla.GlyAlaValGlyLeuGluArqArqAspIrpSerProSe 46
          :::::|||||| |||          |||:::|||
      180 GGAGCGCCGCGTGGTCATCGTTGTCTTTCTCGG.................. 212

       46 rProProAlaIhrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrP 63
           |.|||||||   |||       ||||||
      213 .CCTCCTGCTGGACCT...CCTGGCCTTCAC.................... 239

       63 heAspMetLeuTyrProGluAspSerSerIrpAlaAlaLysAlaProGly 79
                                          ||||||   ||||||:::
      240 ................................GCTGCTGCCGCCCTGCT 257

       80 AlaSerSerArqGluGlnProProGlnGluProGluGlnCysProValIl 96
          |||   :::                    |||||||:::   ||||||
      258 GCCCGGGCT.............GTTGGAGAGCCACGGCCGTGCCCACGACC 295

       96 eAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuI 113
                             ||||||                |||||||||:
      296 CCCTCIATGGCTCCTGGCAGGGCGG..............GGTGGACTGG 330

      113 hrLeuGluGlaHis.........SerLeuGluGlnValGln.SerMet.. 129
          ::   :::::::::          ||   |||:::||||| ||:::
      331 TTTGCCAGCGCAIGCGCATGCCACTGGAGAAGAGCTACAACACTGCCT 380

      126 .........ValValGlyGluValLeuLysAspIleGluThrAlaCys. 138
                  ::::::|||:::::::::::::   ::::::      ||
      381 GTTCGGAGGICTCATTGGCTCGGCATTCTCTGTCCTGCAGTTTCTGTGTG 430

      139 ..LysLeuLeuAsnIleThrAlaAspProMetAspIrpSerProSerAsn 154
            |||   :::   |||:::|||   :::          |||
      431 CGCCACTCACTGGGGCCACCTCTGACTGCTTGGGGAGGCGCCCGGTGATG 480

      155 ValGlnLysIrpLeuLeuIrpThrGluHisGlnIyrArqLeuProProM 171
          :::         ::: :::|||      |||:::         ||||||:
      481 CTGCTGTGCCTGATGGTGTGGCCACCTCATATGCAGTCTGGGCCACCTC 530

      171 eIrpIyrLysAlaPheSerIleGlnAlaIrpIyrSerLeuGlnCysAlaM 187
          ::|||   :::          :||  |||   :::
      531 TCGGAGCTTTGCAGCCTTCCTGGCCTCCAGGCTGATTGGGGGCAATGCA 580
```

```
     188 GluGluGlnPheArgGluArgSerPro.....LeuGlyGlyAspValLe 202
         :::          :::     |||      |||            ||
     581 AAGGGAACGTCAGCCTCTCCACGGCCATCGTGGCTGACCTGGGCTCGCCT 630

     202 uHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgI 219
         |        :::      ||||:|||   :::|||   ::::::   :
     631 CTGGCCCGCAGTCAAGGCATGGCGGTCATTGGGGTGGCCTTCTCACTGGG 680

     219 hrSerProGlyAlaIle.............................His 225
         ::||||||   |||:::                             |||
     681 CTTCACCCTGGGCCCTATGCTCGGAGCCTCCCTGCCCCTGGAAATGGCAC 730

     226 TyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluValAspSe 242
                   ||||:::::      |||     :::   :::::
     731 CCTGGTTTGCCCTGCTCTTCGCAGCCTCCGACCTGCTGTTCATCTTCTGC 780

     242 rSerCys.....................SerGlyGlnProIleHisLeuI 252
         :||||||                    :::|||:::|||:::      |
     781 TTCCTGCCAGAGACGCTGCCCCTGGAGAAACGGGCGCCCTCTATCGCCCT 830

     252 rpGlnPheLeuLysGluLeuLeu 259
         ||     :::   :::|||:::
     841 GGGGTTCCGTGATGCGGCTGATC 863

Seq_name: /cgn2_6/ptodata/1/iia/5A_COMB.seq:us-08-173-508-3

Seq_documentation_block:
; Sequence 3, Application US/08173508
; Patent No. 5616485
;  GENERAL INFORMATION:
;     APPLICANT:  Bartfeld, Daniel
;     APPLICANT:  Butler, Michael J.
;     APPLICANT:  Hadary, Dany
;     APPLICANT:  Jenish, David
;     APPLICANT:  Krieger, Timothy
;     TITLE OF INVENTION:  STREPTOMYCES PROTEASES AND IMPROVED
;     TITLE OF INVENTION:  STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
;     TITLE OF INVENTION:  POLYPEPTIDES
;     NUMBER OF SEQUENCES:  13
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, N.W.
;       CITY:  Washington, D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/173,508
;       FILING DATE:  23-DEC-1993
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  10711-125-228
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-672-5300
;       TELEFAX:  202-672-5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2185 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  531..2069
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  531..902
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  903..2069
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  531..533
;       OTHER INFORMATION:  /note= "Met at position -124
;       OTHER INFORMATION:  represents fMet"
US-08-173-508-3


alignment_scores:
            Quality:   89.50              Length:     80
              Ratio:   1.865               Gaps:      5
 Percent Similarity:   60.000   Percent Identity:  47.500

alignment block:
 US-09-126-945B-2 x US-08-173-508-3/rev   ..

 Align_Seq 1/1  to reverse of: US-08-173-508-3  from: 1  to: 2185

       19 ProProAspThrValSerArg.....ThrGlyLeuGluLysAlaAlaA 33
          ||||||  |||:::|||:::      |||||||:::||| :::|||
      424 CCGCCGAGGACGACTTCGGAAAGTGCCACGGGTGTGAACCGGGCCCCGC 274

       34 laGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProSerPro 49
          |||||||||:::|||:::|||:::          :::||||||||||
      273 GGGGCGCCGTGTCGTTGGACAGGGAGCCCGAGCACGATCAGCCCCCTGCG 224

       50 ...ThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMe 65
             ||||||:::::      |||:::          ||||||
      223 AGAACACCGAGGTCCAAGCCGTCATGT.....CCGTCGTAC........ 188

       65 tGlnTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerS 82
                  |||:::::   |||        ||||||      |||:
      187 .........GAGGAGACTCCGTGGCCACCCGACGCACCGCCCCCTCCA 148

       82 erArgGluGluProProGluGluProGluGluCysPro 94
          :::::|||   |||   :::|||   ::::::   |||
      147 AAGAGGAGCCCCCTCGCAAGAGGAGCACCGCCTCCCC 110

Seq_name: /cgn2_6/ptodata/1/iia/5A_COMB.seq:us-08-265-410-4

Seq_documentation_block:
; Sequence 4, Application US/08265410
; Patent No. 5856166
;  GENERAL INFORMATION:
;     APPLICANT:  Bartfeld, Daniel
;     APPLICANT:  Butler, Michael J.
;     APPLICANT:  Hadary, Dany
;     APPLICANT:  Jenish, David
;     APPLICANT:  Krieger, Timothy
;     APPLICANT:  Malek, Lawrence T.
;     APPLICANT:  Soostmeyer, Gisela
;     APPLICANT:  Walczyk, Eva
;     APPLICANT:  Kryssman, Phyllis
;     APPLICANT:  Jarcen, Shelia
;     TITLE OF INVENTION:  STREPTOMYCES PROTEASES AND IMPROVED
;     TITLE OF INVENTION:  STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
;     TITLE OF INVENTION:  POLYPEPTIDES
;     NUMBER OF SEQUENCES:  21
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, N.W.
;       CITY:  Washington, D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/265,310
;       FILING DATE:  24-JUN-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/173,508
;       FILING DATE:  23-DEC-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  190740/0140
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202 672 5300
;       TELEFAX:  202 672 5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2185 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  531..2069
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  531..902
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  903..2069
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  531..533
;       OTHER INFORMATION:  /note= "Met at position  124
;       OTHER INFORMATION:  represents fMet"
US-08-265-310-3


alignment_scores:
            Quality:   89.50              Length:     80
              Ratio:  1.865                Gaps:      5
 Percent Similarity:  60.000   Percent Identity: 37.500

alignment_block:
 US-09-126-945B-2 x US-08-265-310-3/rev   ..

 Align seq 1/1  to reverse of: US-08-265-310-3  from: 1  to: 2185

       19 ProProAspThrValSerArg.....ThrGlyLeuGlu.LysAlaAlaA 33
          ||| ||   |||:::|||:::      |||||||:::||| :::|||
      323 CCGCCGAGGACGACTTCGGAAAGTGCCACGGGTGTTGAACCGGGCCCCCGC 274

       33 laGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAla 49
          |||||||||:::|||:::|||:::           :::|||||||||
      273 GGGCGCCGTGTCGTCGGACAGGGAGCCCGAGCACGATCAGCCCCCTGCG 224

       50 ...ThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMe 93
             |||||:::::      |||:::        ||||||
      223 AGAACACCGAGGTCCAAGCCGTCATGT.....CCGTCGTAC........ 188

       65 rLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerS 82
                |||::::::   |||          ||||||       |||:
      187 .........GAGGAGACTCCGTGGCCACCCGACCGACCGCCCCCTCCA 148

       82 erArgGluGluProProGluGluProGluGlnCysPro 94
          :::::|||   |||   :::|||   ::::::   |||
      147 AAGAGGAGCCCCCCTCGCAAGAGGAGCCACCGCCTCCCC 110

seq_name: /cgn2_6/prodata/1/inq/6A_COMB.seq:US-08-951-742-3

seq_documentation_block:
; Sequence 3, Application US/08951742
; Patent No. 6127144
;  GENERAL INFORMATION:
;    APPLICANT:  Bartfeld, Daniel
;    APPLICANT:  Michael J. Butler
;    APPLICANT:  Dany Hadary
;    APPLICANT:  David Jenish
;    APPLICANT:  Tim Krieger
;    APPLICANT:  Lawrence T. Malek
;    APPLICANT:  Gisela Stollmeyer
;    APPLICANT:  Eva Walczyk
;    APPLICANT:  Phyllis Kryqsman
;    APPLICANT:  Sheila Garven
;    TITLE OF INVENTION:  METHOD FOR EXPRESSION OF PROTEINS IN
;    TITLE OF INVENTION:  BACTERIAL HOST CELLS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/951,742
;      FILING DATE:  16-OCT-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bent, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  190740/0140
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2185 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  531..2066
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  531..902
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  903..2066
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  531..533
;      OTHER INFORMATION:  /product= "Met at position -124
;      OTHER INFORMATION:  represents fMet"
US-08-951-742-3


alignment_scores:
            Quality:   89.50              Length:     80
              Ratio:  1.865                Gaps:      5
 Percent Similarity:  60.000   Percent Identity: 37.500

alignment_block:
 US-09-126-945B-2 x US-08-951-742-3/rev   ..

 Align seq 1/1  to reverse of: US-08-951-742-3  from: 1  to: 2185

       19 ProProAspThrValSerArg.....ThrGlyLeuGlu.LysAlaAlaA 33
```

```
          ||||||   |||:::|||:::      |||||||:::||| :::|||
       823 CCGCCGAGGACGACTTCGGAAAGTGCCACGGGTGTTGAACCGGGGCCCCGC 274

        33 laGlyAlaValGlyLeuGluArqArqAspTrpSerProSerProProAla 49
           ||||||||||:::|||:::|||:::            :::||||||||||
       273 GGGGCGCCGTGTCGTTGGACAGGGAGCCCGAGCACGATCAGCCCCCTGCC 224

        50 ...ThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMe 65
              ||||||:::::      |||:::      ||||||
       223 AGAACACCGAGGTCCAACGCCCTCATGT.....CCGTCCTAC........ 188

        65 rLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerS 82
                  |||::::::::  |||      ||||||      |||:
       187 .........GAGGAGACTCCGTGGCCACCCGACGCACCGCCCCCTCCA 148

        82 erArqGluGluProProGluGluProGluGluCysPro 94
           :::::|||   |||   :::|||   ::::::   |||
       147 AAGAGGAGCCCCCCTCGCAAGAGGAGCACCGCCTCCCC 110

seq name: /cgn2_6/ptodata/1/ina/68_COMB_seq/US-09-141-212-5

seq documentation block:
; Sequence 5, Application US/09141212
; Patent No. 6200777
;  GENERAL INFORMATION:
;    APPLICANT:  MACPHEE, COLIN
;    APPLICANT:  PATEL, LISA
;    TITLE OF INVENTION:  NOVEL COMPOUNDS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ratner & Prestia
;      STREET:  P.O. Box 980
;      CITY:  Valley Forge
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/141,212
;      FILING DATE:  27-AUG-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 97306807.5
;      FILING DATE:  01-SEP-1997
;      APPLICATION NUMBER:  EP 98300687.5
;      FILING DATE:  30-JAN-1998
;      APPLICATION NUMBER:  GB 9807720.9
;      FILING DATE:  08-APR-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Prestia, Paul F
;      REGISTRATION NUMBER:  24,031
;      REFERENCE/DOCKET NUMBER:  GP-30012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0700
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2693 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-141-212-5


alignment_scores:
            Quality:    89.50          Length:     330
              Ratio:    0.640            Gaps:      13

 Percent Similarity:  43.030   Percent Identity:  20.606

alignment block:
 US-09-126-945B-2 x US-09-141-212-5   ..

 Align_seq 1/1  to: US-09-141-212-5  from: 1  to: 2693

        6 ProGlyLeuSerSerValSerPro.....SerHisLeuLeuLeuPr 19
          |||||   :::|||:::::::|||          |||   |||||
     1402 CCTGGACGAACCAGTATCACCCCCTTCCCGGGCCCAGCGCTCCCGCTCCC 1451

       19 oProAspThrValSerArqThrGlyLeuGluLysAlaAlaAlaGlyAlaV 36
          |    :::   :::|||       :::   |||:::|||
     1452 TGCCCCAGCCCAAACTCGGTACCCAGTGCCCAGCTGCCTTCTGGCCCCT 1501

       36 alGlyLeuGluArqArqAspTrpSerProSerProProAlaThrProGlu 52
                                    |||:::||||||      ||||||
     1502 GCTTCACG.............CCCCCAGCGGCCGCCGCCCCTTCCTGAG 1536

       53 Glu.....GlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
                |||   :::::|||:::   :::          :::
     1537 TGGAGATGAGGATCCCAAGGCTTCCACGCTACGTGTTGTGGTCTTTGGCT 1586

       67 rProGluAsp.........SerSerTrpAlaAlaLysAlaProGlyAlaS 81
             ::|||          ::::::   :::::::::   |||:::|||:
     1587 CCGATCGGATTCAGGGAAGGTGGCTCGGGCGTACAGGAACCTTCGGCGG 1636

       81 erSerArqGluGluPro..................................... 86
          ::::::::::::::|||
     1637 CTGGAGAACAATCGCCACTCCTCACACGGTTCTTCAAACTTCACTTCTT 1686

       86 ..... 86
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
     1687 CTACGTGCCTGTGAAGCGAAGTCATGGGACCAGCCCTGGTGCCTGTCCAC 1736

       87 .ProGluGluProGlu.GlnCysProValIleAspSerGluAlaProAla 102
          |||:::|||||:::      |||      ||||||      |||
     1737 CCCCTCGGAGCCAGACGCCCTCACCCCCGACAGACTCCCCTAGGCACGCC 1786

      103 GlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLe 119
          :::      :::|||          |||   |||      ||||:|||
     1787 AGCCCTGCTGAGCTG.........GGCACCACCCCCATGGGAGGAGAGCAC 1827

      119 uGluGlnVal.................................GluSerM 125
          :::::::::                                  :::
     1828 CAATGACATCTCCCACTACCTCGGCATGCTGGACCCCTGGTATGAGCGCA 1877

      125 etValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeu 141
          |||:::|||   :::          |||:::   ||||:::   |||
     1878 ATGTACTGGGGCTCATGCACCTGCCCCCTGAAGTCCTGTGCCAGTCCCTG 1927

      142 AsnIleThrAlaAspProMetAspTrpSerProSerAsnVal.GluLysT 158
          :::      ||||:::   ::::::   |||||||||||||:    :::|
     1928 AAGGCTGAAGCCCAGGCCCTGGAGGGCTCCCCAACCCAGCTGCCCATCCT 1977

      158 rpLeuLeuTrpThrGluHisGlnTyrArqLeuProPro.MetGlyLysAl 174
          |||||      :::          |||||||||   |||   ||
     1978 GGCTGACATGCTACTCTACTACTGCCGCTTTGCCGCCAGACCGGTGCTGC 2027

      174 aPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnP 191
          |   :::                 :::|||          :::|||:::::
     2028 TGCAAGTCTATCAGACCGAGCTGACCTTCATCACTGGGGAGAAGACGACA 2077

      191 heArqGluArqSerProLeuGlyGlyAspValLeuHisAlaHis..... 205
                :::      |||||||:::      |||:::::|||
     2078 GAGATCTTCATCCACTCCTTGGAGCTGGGTCACTCCGCTGCCACACGTGC 2127

      206 .........LeuAspIleTrpLysSerAlaAla.................. 218
                  :::      |||:::::::|||||
     2128 CATCAAGGCGTCAGGTCCTGGCAGCAAGCGGCTGGGCATCGATGGCGACC 2177

      218 ..... 218
```

```
     2178 GGGAGGCTGTTCCTCTAACACTACAGATTATTTACAGCCAGGGGGCCATC 2227

      214 ..TrpMetLysGluArgThrSerProGlyAlaIleHisTyr.CysAlaS 259
             |||   |||:::      :::||||||          |||    :::::
     2228 AGTGGACCAAGTCGCTGGAGCAACCTGGAGAAGGTCTGTACCTCCGTGAA 2277

      259 rThrSerGluGluSerTrpThrAspSerGluValAspSerSerCysSerG 246
          :   ::::::            ::::::|||:::|||||||||   :::: 
     2278 CCTTAACAAGGCCTGCCGGAAGCAGGAGGAGCTGGATTCCAGCATGGAGC 2327

      246 lyGlnProIleHisLeuTrpGlnPheLeuLys 256
          ::      ::::::|||   :::   ::::|| 
     2328 CGCTGACGCTAAACCTGACAGAAGTGGTGAAA 2359

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:us-08-827-208-2

seq_documentation_block:
; Sequence 2, Application US/08827208
; Patent No. 6025178
;  GENERAL INFORMATION:
;     APPLICANT:  Chiou, Xue-Chiou C.
;     APPLICANT:  Kramer, Ruth M.
;     APPLICANT:  Pickard, Richard T.
;     APPLICANT:  Sharp, John D.
;     APPLICANT:  Stuttler, Beth A.
;     TITLE OF INVENTION:  HUMAN PHOSPHOLIPASE A2 AND RELATED
;     TITLE OF INVENTION:  NUCLEIC ACID COMPOUNDS
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Eli Lilly and Company
;       STREET:  Lilly Corporate Center
;       CITY:  Indianapolis
;       STATE:  Indiana
;       COUNTRY:  United States of America
;       ZIP:  46285
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/827,208
;       FILING DATE:  28-MAR-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/014,608
;       FILING DATE:  29-MAR-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/041,264
;       FILING DATE:  19-MAR-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Gaylo, Paul J.
;       REGISTRATION NUMBER:  36,808
;       REFERENCE/DOCKET NUMBER:  X-10610
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (317) 276-0756
;       TELEFAX:  (317) 276-3861
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3085 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  48..2786
US-08-827-208-2


alignment_scores:
            Quality:   89.50              Length:    436
              Ratio:  0.511                Gaps:     26

 Percent Similarity:  40.138   Percent Identity: 21.789

alignment_block:
 US-09-126-945B-2 x US-08-827-208-2   ..

 Align seq 1/1  to: US-08-827-208-2  from: 1  to: 3085

       4 AlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProPr 20
         |||||||||      ::::||  ::|||                 |||
    1423 GCGICTCCT......ACATCACCGGGGCCTCGGGCTCCACCTGGGCCTTG 1466

      20 oAspThrValSerArgThr......GlyLeuGluLysAlaAlaAlaGlyA 35
           |||       ||||||      ||||||:::::::::       |||
    1467 GCCAACCTTTATGAGGACCCAGACTGGTCTCAGAAGGACCTGGCAGGGCC 1516

      35 laValGlyLeu..............GluArgArgAspTrpSerProSer 46
         ::::::                  |||   ::::||
    1517 CACTGAGTTCCTGAACAGCCCAGTTGACCAAGAACAAGCTGGGTGTGCTGG 1566

      47 ProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPh 63
         ||||||||||:::                      |||
    1567 CCCCCAGCCAGC                      TGCAGCGGTAC 1589

      63 eAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla..... 77
                ::::|||:::    :::|||||||:::::|||
    1590 CGGCAGCAGCTGGCCCAGCCTGCCCGCTTGGGCTACCCAAGCTGCTTCAC 1639

      78 ..........ProGlyAlaSerSerArg.........GluGluProPro 87
                    |||:::::::::   |||          ::::||
    1640 CAACCTGTGGGCCCTCATCAACGAGGCGCTGCTGCATGATGAGCCCCATG 1689

      88 GluGluProGluGln..................................... 92
                 ::|||
    1690 ATCACAAGCTCTCAGATCAACGGGAGGCCCTGAGTCATGGCCAGAACCCT 1739

      93 .CysProValIleAspSerGlnAlaProAlaGlySer.............L 105
         ||||||           ::::::|||   |||                 |
    1740 CTGCCCATCTACTGTGCCCTCAACACCAAAGGCCAGAGCCTGACCACTTT 1789

     105 euAspLeuValProGlyGlyLeuThrLeuGluGluHisSer LeuGluGl 121
         ||:::|||     ||||::   ::|||      ::::|| |||   ||
    1790 TGAATTGGGCAGTGGTGGAGTTCTCTGCCTACGAGGTGGGCTTCCCA 1839

     121 nValGlnSerMet.Val.  .                    Val 127
         ||||:::::::::: :::                     :::
    1840 AGTACGGGGCCTTCATGCCCTCTGACCTCTTTGGCTCCGAGTTCTTTATG 1889

     128 GlyGluValLeuLysAspIleGluThrAla......CysLysLeuLeuAs 142
         |||:::::::|||   :::     :::      |||   |||   ::
    1890 GGCCAGCTCATCAACAGCCTTCCTGACTCCCCCATCTGCTTCTTAGAAGG 1939

     142 nIleThrAla.......................AspProMetAspTrpS 151
         :|||   :::                     |||   :::   |||:
    1940 TATCTGCAGCAACCTGTATGCAGCCAACCTCCAGGACACCTTATACTGGG 1989

     151 erProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArg 167
         ::   ...   |||   |||          |||.........|||   :::
    1990 CCTCAGAGCCCAGCCAGTTCTGGGACCGCTGGGTCAGGAACCAGCCAAC 2039

     168 Leu.....................................ProPro..... 170
         |||                                    ||||||
    2040 CTGGACAAGCAGCAGGTCCCCCTTCTGAAGATACAACAACCACCCTCAAC 2089

     171 .........MetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuC 184
                  :::::::::   |||   ::|||         :::
    2090 AGCCCCAGAAIAGCTGACTTTTTCACCCATCTTCTGACCTGGGGTCCAC 2139

     184 ysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAsp 200
           |||::::::   ::::::|||   :::          :::      |||
    2140 TGGCCCAGGCCACACATAATTTCCTGCGTGGCCTCCATTTCCACAAAGAC 2189

     201 ValLeu...HisAlaHisLeuAspIleTrpLysSerAlaAla.TrpMetL 216
           :::   |||   |||::::::   ||||||:::::::: ||||||
```

```
  2190 TACTTTCAGCATCCTCACTTCTCCACATGGAAAGCTACCACTCTGGATGG 2239

   216 ysGluArgThrSer..................................... 220
         :::   ||||||
  2240 GCTCCCCAACCAGCTGACACCCTCGGAGCCCCACCTGTGCCTGCTGGATG 2289

   221 ....................ProGlyAlaIleHisTyrCysAlaSer..... 229
                           |||:::|||         |||:::
  2290 TTGCCTACCTCATCAATACCAGCTGCCTGCCCCTCCTGCAGCCCACTCCC 2339

   230 .......ThrSerGluGluSerTrp......ThrAspSerGluValAspS 242
             |||||||:::      |||      |||:::::::|||   :::|
  2340 CACGTGGACCTCATCCTGTCATTGGACTACAACCTCCACGGAGCCTCCCA 2389

   242 erSerCysSerGly.GlnProIle........................... 249
       |||||||||||:::    |||:::
  2390 GCAGTTGCAGCTCCTGGGCCGGTTCTGCCAGGAGCAGGGGATCCCGTTCC 2439

   250 ....HisLeuTrpGlnPheLeuLysGluLeu.................... 258
           ||||||   |||      :::
  2440 CACCCATCTGCCCAGCCCCGAAGAGCAGCTCCAGCCTGGGAGTGCCAC 2489

   259 ..LeuLeuLysProHis.SerTyrGlyArgPheIleArgTrpLeuAsnLy 274
         ||||||:::||||||      |||
  2490 ACTTTCTCCGACCCCACCTGCCCCGGA....................... 2516

   274 sGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLe 291

  2516 ..................................................... 2516

   291 rpGlyIleArgLysAsnArgProAla...MetAsnTyrAspLysLeuSer 306
                              ||||||   :::::::::      ::::||
  2517 ..................GCCCCTGCGGTGCTGCACTTTCCTCTGGTCAGC 2549

   307 ArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAsp 322
         |||:::||:::::||          |||:::   |||   |||:::
  2550 GACTCCTTCCGGGAGTACTCGGCCCCTGGGGTCCGGCGGACACCCGAG 2597

seq name: /cgn2_6/ptodata/1/iga/6A_COMB.seq:US_08_827_208_4

Seq_docOumentation block:
; Sequence 4, Application US/08827208
; Patent No. 6025178
;  GENERAL INFORMATION:
;    APPLICANT:  Chiou, Xue Chiou C.
;    APPLICANT:  Kramer, Ruth M.
;    APPLICANT:  Pickard, Richard T.
;    APPLICANT:  Sharp, John D.
;    APPLICANT:  Striffler, Beth A.
;    TITLE OF INVENTION:  HUMAN PHOSPHOLIPASE A2 AND RELATED
;    TITLE OF INVENTION:  NUCLEIC ACID COMPOUNDS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/827,208
;      FILING DATE:  28 MAR 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/014,608
;      FILING DATE:  29 MAR 1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/041,264
;      FILING DATE:  19 MAR 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-10610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276 0756
;      TELEFAX:  (317) 276 3861
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3085 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  mRNA
US 08-827 208 4


alignment_scores:
            Quality:   89.50              Length:    436
              Ratio:   0.511               Gaps:     26
 Percent Similarity:  40.148   Percent Identity: 21.789

alignment_block:
 US 09-126-945B 2 x US 08-827 208-4    ..

 Align_Seq 1/1  to: US 08 827 208 4  from: 1  to: 3085

      4 AlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProPr 20
        |||||||||       :::|||   :::|||                |||
   1423 GCGGCUCCU......ACAUCACCGGGGCCUCCGGGCUCCCACCUGGGCCUGG 1466

     20 aAspThrValSerArgThr......GlyLeuGluLysAlaAlaAlaGlyA 35
        |||      ||||||      ||||||:::::::::      |||
   1467 GCCAACCUUUAUGAGGACCCAGACUGGGUCUCAGAAGGACCUGGCAGGGCC 1516

     35 IaValGlyLeu..............GluArgArgAspTrpSerProSer 46
        ::::::                    |||   :::|||
   1517 CACUGAGUUGCUGAAGACCCAGGUGACCAAGAACAAGCUGGGUUGGCUGG 1566

     47 ProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPh 63
        ||||||||||:::                          |||
   1567 CCCCCAGCCAGC.........................UGCAGCGGUAC 1589

     63 aAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla....... 77
                  :::|||:::      :::||||||:::::|||
   1590 CGGCAGGAGCUGGCCGAGCGGCCCCGCUUGGGCUACCCAAGCUGCUUCAC 1639

     78 ..........ProGlyAlaSerSerArg.........GluGluProPro 87
                 |||:::::::::::   |||             :::|||
   1640 CAACCUGUGGGGCCCUCAUCAACGAGGCGCUGCUGCAUGAUGAGCCCCAUG 1689

     88 GluGluProGluGln..................................... 92
                ::::|||
   1690 AUCACAAGCUCUCAGAUCAACGGGAGGCCCUGAGUCAUGGCCAGAACCCU 1739

     93 .CysProValIleAspSerGlnAlaProAlaGlySer............L 105
         ||||||          :::::|||   |||              |
   1740 CUGCCCAUCUACUGUGCCCUCAACACCAAAGGGCAGAGCCUGACCACUUU 1789

    105 euAspLeuValProGlyGlyLeuThrLeuGluGluHisSer.LeuGluGl 121
        ||:::|||      |||:::   :::|||      :::||| |||   ||
   1790 UGAAUUUGGGGAGUGGUGCGAGUUCUCUCCCUACGAGCUCGGCUUUCCCCA 1839

    121 nValGluSerMet.Val.............................Val 127
        ||||:::::::::: :::                            :::
   1840 AGUACGGGGCCUUCAUCCCUCUGAGCUCUUUGGCUCCGAGUUCUUUAUG 1889

    128 glyGluValLeuLysAspIleGluThrAla......CysLysLeuLeuAs 142
        |||:::::::::|||   :::      :::      |||   |||   ::
   1890 GGGCAGCUGAUGAAGAGCCUUCCUGAGUCCCGCAUCUGCUUCUUAGAAGG 1939

    142 nIleThrAla.......................AspProMetAspTrpS 151
```

```
      :|||   :::                                  |||   :::   |||:
  1940 UAUCUGGCAGCAACCCUGUAUGCAGCCAACCCUCCAGGACAGCCUUAUACUGGG 1989

       151 erProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArg 167
           ::   :::      |||   |||      |||::::::::::|||   :::
  1990 CCUCAGACCCCAGCCAGUUCUGGGACCGCUGGCUCACGAACCAGGCCAAC 2039

       168 Leu                                  ProPro      170
           |||                                  ||||||
  2040 CUGGACAAGGAGCAGGUCCCCCUUCUGAAGAUAGAAGAACCACCCUCAAC 2089

       171 ......... MetGlyLysAlaPheGlnGluLeuAlaGlyLysGlnLeuG 191
                   :....  |||    |||
  2090 AGCCGGCAGAAUAGCUGAGUUUUUCACCGAUCUUCUGACGUUGGCGUCCAC 2139

       184 ysAlaMetSerGluGluPheArgGluArgSerProLeuGlyGlyAsp 200
             |||::::::   ::::::|||   :::          :::     |||
  2140 UGCCCCAGCCCAGACAUAACUCCUCGCGUCGCCUCCAUUCCACAAACAC 2189

       201 ValLeu...HisAlaHisLeuAspIleTrpLysSerAlaAla.TrpMetL 216
           :::   |||   |||::::::   ||||||:::::::: ||||||
  2190 UACUUUCAGCAUCCUCACUUCUCCACAUGGAAAGCUACCACUCUGGAUGG 2239

       216 ysGluArgThrSer..................................... 220
             :::   ||||||
  2240 GCUCCCCAACCACCUGACACCCUCGCAGCCCCACCUCUCCCUGCUGGAUG 2289

       221 ..............ProGlyAlaIleHisTyrCysAlaSer..... 229
                        ||::::|||         ||:::
  2290 UUGGCUACCUCAUCAAUACCAGCUGCCUGCCCCUCCUGCAGCCCACUGG 2339

       230 .......ThrSerGluGluSerTrp......ThrAspSerGluValAspS 242
                 ||||||:::      |||      |||::::::|||   :::|
  2340 GACCUGGACCUCAUCCUGUCAUUGGACUACAACCUCCACGGAGCCUUCCA 2389

       242 erSerCysSerGly.GlnProIle........................... 249
           ||||||||||:::   ||:::
  2390 GCAGUUGCAGCUCCUGGGCCGGUUCUGCCAGGAGCAGGGGAUCCCGUUCC 2439

       250 ....HisLeuTrpGlnPheLeuLysGlnLeu.................... 258
               ||||||   |||      :::
  2440 CACCCAUCUGCCCAGCCCCGAAGAGCAGCUCCAGCCUCGGGAGUGCCAC 2489

       259 ..LeuLeuLysProHis.SerTyrGlyArgPheIleArgTrpLeuAsnLy 274
             |||||||:::||||||       |||
  2490 ACCUUCUCCGACCCCACCUGCCCCGGA....................... 2516

       274 sGlnLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuT 291

  2516 ..................................................... 2516

       291 rpGlyIleArgLysAsnArgProAla...MetAsnTyrAspLysLeuSer 306
                                |||||| ::::::::      :::|||
  2517 .........        GCCCCUGCCGGUGCUGCCACUUUCCUCUGGUCAGC 2549

       307 ArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAsp 322
           |||:::|||:::|||          |||:::  |||  |||:::
  2550 GACUCCUUCCGGGACUACUCGCCCCUGGGGUCCGGCGACACCCGAG 2597

seq_type: /tgsi2_f/prokaria/1/ftn/96_20MB seq:US-09-500-358-2

seq_documentation_block:
; Sequence 2, Application US/09500358
; Patent No. 6197569
;  GENERAL INFORMATION:
;     APPLICANT:  Chiou, Xue-Chiou C.
;     APPLICANT:  Kramer, Ruth M.
;     APPLICANT:  Pickard, Richard T.
;     APPLICANT:  Sharp, John D.
;     APPLICANT:  Strifler, Beth A.
;     TITLE OF INVENTION:  HUMAN PHOSPHOLIPASE A2 AND RELATED
;     TITLE OF INVENTION:  NUCLEIC ACID COMPOUNDS
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Eli Lilly and Company
;       STREET:  Lilly Corporate Center
;       CITY:  Indianapolis
;       STATE:  Indiana
;       COUNTRY:  United States of America
;       ZIP:  46285
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/500,358
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/827,208
;       FILING DATE:  28-MAR-1997
;       APPLICATION NUMBER:  US 60/014,608
;       FILING DATE:  29-MAR-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/041,264
;       FILING DATE:  19-MAR-1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Gaylo, Paul J.
;       REGISTRATION NUMBER:  36,808
;       REFERENCE/DOCKET NUMBER:  X-10610
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (317) 276-0756
;       TELEFAX:  (317) 276-3861
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3085 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  48..2786
US-09-500-358-2


alignment_scores:
            Quality:   89.50              Length:    436
              Ratio:  0.511                Gaps:     26
 Percent Similarity:  40.138   Percent Identity: 21.789

alignment_block:
 US-09-126-945B-2 x US-09-500-358-2     .

 Align seq 1/1  to: US-09-500-358-2  from: 1  to: 3085

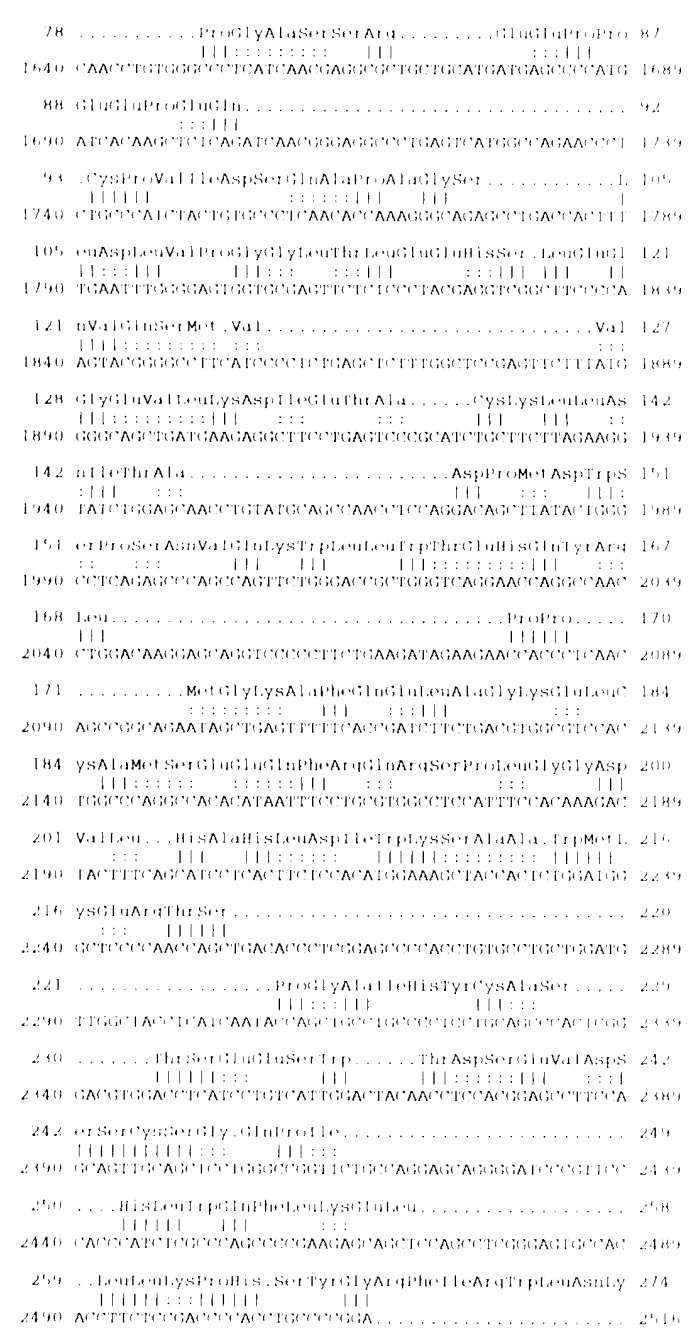
       4 AlaSerProGlyLeuSerSerValSerProSerHisLeuLeuProPr 20
         |||||||||      :::|||   :::|||                |||
    1423 GCGCCCCCC.....ACACACCGGGGCCUGGGCCUCCACCUGGGCCUUC 1466

      20 oAspThrValSerArgThr......GlyLeuGluLysAlaAlaAlaGlyA 35
           |||      ||||||      ||||||:::::::::      |||
    1467 GCCAACCTTTATGAGGACCCACAGUGGUCUCAGAAGGACCUGGCAGGGCC 1516

      35 laValGlyLeu                GluArgArgAspTrpSerProSer 46
         :::::                      |||   :::|||
    1517 CACUGAGUUGCUGAAGACCCAGGUGACCAAGAACAAGCUGGGUGUGCUGG 1566

      47 ProProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPh 63
         ||||||||||:::                          |||
    1567 CCCCCAGCCAGC                        UGCAGCGGUAC 1589

      63 eAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAla..... 77
               :::|||:::      :::|||||||:::::|||
    1590 CGGCAGGAGCUGGCCGAGCCGCCCGCUUGGGCUACCCAAGCUGCUUCAC 1639
```


78 ....ProGlyAlaSerSerArg.....GluGluProPro 87
1640 CAACCTGTGGGCCCTCATCAACGAGGCGCTGCTGCATGATGAGCCCCATC 1689
88 GluGluProGluGln..... 92
1690 ATCACAAGCTCTCAGATCAACGGGAGGCCCTGAGTCATGGCCAGAACCCT 1739
93 CysProValIleAspSerGluAlaProAlaGlySer.....L 105
1740 CTGCCCATCTACTGTGCCCTCAACACCAAAGGGCAGAGCCTGACCACTTT 1789
105 euAspLeuValProGlyGlyLeuThrLeuGluGluHisSer.LeuGluGl 121
1790 TGAATTTGGGGAGTGGTGCGAGTTCTCTCCCTACGAGGTCGGCTTCCCCA 1839
121 nValGlnSerMet.Val.....Val 127
1840 AGTACGGGGCCTTCATCCCCTCTGAGCTCTTTGGCTCCGAGTTCTTTATG 1889
128 GlyGluValLeuLysAspIleGluThrAla.....CysLysLeuLeuAs 142
1890 GGGCAGCTGATGAAGAGGCTTCCTGAGTCCCGCATCTGCTTCTTAGAAGG 1939
142 nIleThrAla.....AspProMetAspTrpS 151
1940 TATCTGGAGCAACCTGTATGCAGCCAACCTCCAGGACAGCTTATACTGGG 1989
151 erProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGluTyrArg 167
1990 CCTCAGAGCCCAGCCAGTTCTGGGACCGCTGGGTCAGGAACCAGGCCAAC 2039
168 Leu.....ProPro..... 170
2040 CTGGACAAGGACGAGTCCCCCTTCTGAAGATAGAAGAACCACCCTCAAC 2089
171 .....MetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuC 184
2090 AGCCGGCAGAATAGCTGAGTTTTCACCGATCTTCTGACGTGGCGTCCAC 2139
184 ysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAsp 200
2140 TGGCCCAGGCCACACATAATTTCCTGCGTGGCCTCCATTTCCACAAAGAC 2189
201 ValLeu...HisAlaHisLeuAspIleTrpLysSerAlaAla.TrpMetL 216
2190 TACTTTCAGCATCCTCACTTCTCCACATGGAAAGCTACCACTCTGGATGG 2239
216 ysGluArgThrSer..... 220
2240 GCTCCCCAACCAGCTGACACCCTCGGAGCCCCACCTGTGCCTGCTGGATG 2289
221 .....ProGlyAlaIleHisTyrCysAlaSer..... 229
2290 TTGGCTACCTCATCAATACCAGCTGCCTGCCCCTCCTGCAGCCCACTCGC 2339
230 .....ThrSerGluGluSerTrp.....ThrAspSerGluValAspS 242
2340 GACGTGGACCTCATCCTGTCATTGGACTACAACCTCCACGGAGCCTTCCA 2389
242 erSerCysSerGly.GlnProIle..... 249
2390 GCAGTTGCAGCTCCTGGGCCGGTTCTGCCAGGAGCAGGGATCCCGTTCC 2439
250 ....HisLeuTrpGlnPheLeuLysGluLeu..... 258
2440 CACCCATCTCGCCCAGCCCCGAAGACAGCTCCACCCTCGGCAGTGCCAC 2489
259 ..LeuLeuLysProHis.SerTyrGlyArgPheIleArgTrpLeuAsnLy 274
2490 ACTTTCTCCGACCCCACCTGCCCCGGA..... 2516


274 sGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuT 291
2516 2516
291 rpGlyIleArgLysAsnArgProAla...MetAsnTyrAspLysLeuSer 306
2517 .....GCCCCTGCGGTGCTGCACTTTCCTCTGGTCAGC 2549
307 ArgSerIleArgAlaTyrTyrLysLysGlyIleIleArgLysProAsp 322
2550 GACTCCTTCCGGGAGTACCGGCCCCTGGGGTCCGGCGGACACCCGAG 2597

```
OM of: US-09-126-945B-2 to: GenEmbl:*   out_format : pts

Date: Sep 18, 2001  7:08 AM

About: Results were produced by the GenPept software, version 1.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL frame+_p2n.model -DEV xlp
  -Q /tmp/_2/US-09-126-945B-2/format_17_32601_1207_3545/app_query.fasta_1 -DB
   DB GenEmbl -QFMT=fastap  SUFFIX rqc  GAPOP 12.000  GAPEXT 4.000
  -MINMATCH=0.100 -LOOPCL 0.000  -LOOPEXT=0.000  QGAPOP=4.500
   QGAPEXT 0.050  -XGAPOP 10.000  XGAPEXT=6.500  FGAPOP 6.000
  -FGAPEXT 7.000  -YGAPOP 10.000  YGAPEXT 6.500  DELOP 6.000
   DELEXT 7.000 -START=1 -MATRIX blosum62  TRANS human40.cdi
  -LIST 150  -DOCALIGN=200  -THR_SCORE pct  -THR_MAX=100  -THR_MIN 0
  -ALIGN=100  MODE LOCAL  -OUTFMT pts  -NORM=qal  MINLEN 0
   MAXLEN=2000000000 -USER US-09-126-945B-2_COMB_1_5069  NCPU 4
   DCPU=3  -LONGLOG  -NO_XLPXY  WAIT -THREADS 1

Search Information block:
Query: US-09-126-945B-2
Query length: 435
Database: GenEmbl:*
Database sequences: 1344157
Database length: 856060004
Search time (sec): 1484.020000


score_list:
Sequence                Strd Orig    ZScore    EScore Len   ! Documentation ...
qb_pat1:AX048148          + 1785.00 2385.34  1.3e-124  1891  ! AX048148 Sequence 1 from Paten
qb_pr4:AF071538           + 1785.00 2385.34  1.3e-124  1891  ! AF071538 Homo sapiens Ets tran
qb_pr1:AB031549           + 1785.00 2385.33  1.3e-124  1898  ! AB031549 Homo sapiens PSE mRNA
qb_pat1:AX048159          + 1785.00 2381.10  2.2e-124  3317  ! AX048159 Sequence 12 from Pate
qb_rol:AB019436           + 1581.00 2045.46  1.1e-105  1704  ! AB019436 Mus musculus mRNA for
qb_pr6:AL157272           -  790.50 1018.27  1.8e-48   15060  ! AL157272 Human DNA sequence
qb_in2:AF206964           -  487.50  645.29  1.0e-27   1842  ! AF206964 Strongylocentrotus p
qb_htq4:AC015874          -  371.00  466.00  1.0e-17   38178  ! AC015874 Drosophila melanogas
qb_in1:AC008253           -  371.00  454.21  4.6e-17   180887  ! AC008253 Drosophila melanoga
qb_in1:AC007818           -  371.00  453.90  4.8e-17   188549  ! AC007818 Drosophila melanoga
qb_in1:AC008217           -  371.00  453.71  4.9e-17   193245  ! AC008217 Drosophila melanoga
qb_in2:AE003762           -  371.00  452.36  5.8e-17   230940  ! AE003762 Drosophila melanoga
qb_in3:DRODETS1A          -  369.50  487.24  6.7e-19   362  ! M98474 Drosophila melanogaster
qb_ov:XLA34125            -  307.50  403.14  4.2e-14   2012  ! AJ234125 Xenopus laevis ets ge
qb_pr5:AF147782           -  300.50  396.25  7.8e-14   1447  ! AF147782 Homo sapiens tel rela
qb_pr10:HUMERG2           -  300.50  390.32  1.7e-13   3166  ! M17254 Human erg2 gene encodin
qb_pr4:AF049703           -  300.00  399.20  5.8e-14   899  ! AF049703 Homo sapiens E74-like
qb_pat1:AR103486          +  299.00  392.14  1.3e-13   2280  ! AR103486 Sequence 9 from Patei
qb_pat1:AR103487          +  299.50  390.77  1.6e-13   2498  ! AR103487 Sequence 10 from Pate
qb_pr10:HUMETS2A          +  299.00  390.83  1.6e-13   2269  ! J04102 Human erythroblastosis
qb_pat1:AR103485          +  299.00  390.32  1.7e-13   2428  ! AR103485 Sequence 8 from Patei
qb_in3:CELF22A3           +  298.50  370.59  2.1e-12   29997  ! U11547 Caenorhabditis elegans
qb_ov:XLA34125            -  297.00  388.91  1.8e-13   1768  ! AJ234125 Xenopus laevis ets ge
qb_pr5:AF115403           +  297.00  388.09  2.2e-13   2287  ! AF115403 Homo sapiens Ets tran
qb_pr5:AF115402           +  296.00  386.22  2.8e-13   2452  ! AF115402 Homo sapiens Ets tran
qb_pat2:E31254            +  294.50  388.21  2.2e-13   1447  ! E31254 Protein having cell cul
qb_pr10:HSU13044          +  293.00  383.86  3.8e-13   1970  ! U13044 Human nuclear respirato
qb_pr5:AK027344           +  292.50  384.14  3.7e-13   1736  ! AK027344 Homo sapiens cDNA, FL
qb_pr10:HUMERG11          +  292.50  379.69  6.5e-13   3126  ! M21535 Human erg protein (ets
qb_ov:GGERG               +  292.00  384.50  3.5e-13   1516  ! X77159 G.gallus ERG mRNA. 4/11
qb_pat2:E31254            +  292.00  384.44  3.5e-13   1528  ! E31254 Protein having cell cul
qb_ov:AF257741            +  291.50  382.11  4.9e-13   1992  ! AF257741 Danio rerio ETS1 60
qb_pr9:HSAJ76205          +  290.50  381.76  5.1e-13   1668  ! AJ276205 Homo sapiens mRNA for
qb_ov:DRE249590           +  289.50  381.56  5.1e-13   1436  ! AJ249590 Danio rerio mRNA for
qb_pr10:HUMTEF160         +  289.00  379.26  6.9e-13   1781  ! D13418 Human mRNA for transcri
qb_rol:AF049702           +  288.00  376.23  1.0e-12   2221  ! AF049702 Mus musculus E74 like
qb_vi2:PFAEV1             +  288.00  375.45  1.1e-12   2466  ! X80114 Avian erythroblastosis
qb_om:AF057717            +  288.00  374.30  1.3e-12   1184  ! AF057717 ovis aries transcript
qb_htq4:AC014838          -  281.00  336.78  1.6e-10   117376  ! AC014838 Drosophila melanoga
qb_in1:AC007984           -  281.00  333.75  2.4e-10   174894  ! AC007984 Drosophila melanoga
qb_in2:AE003758           -  281.00  331.84  3.0e-10   225092  ! AE003758 Drosophila melanoga
qb_pr5:AF116509           +  280.50  368.90  2.6e-12   1551  ! AF116509 Homo sapiens Ets 1
qb_pr10:S45205            +  278.50  365.65  4.0e-12   1674  ! S45205 ETS 1 Friend leukemi
qb_pat1:A36461            +  278.50  361.43  6.8e-12   2916  ! A36461 Sequence 2 from Pate
qb_pat1:AR080101          +  278.50  361.38  6.8e-12   2948  ! AR080101 Sequence 3 from pa
qb_pr9:HSBPMFLI           +  278.50  361.38  6.8e-12   2948  ! X67001 H.sapiens HUMFLI 1 m
qb_pr10:HUMERGBFLI        +  278.50  361.33  6.9e-12   2957  ! M98833 Homo sapiens ERGB tr
qb_pr7:BC001670           +  278.50  361.32  6.9e-12   2959  ! BC001670 Homo sapiens, Etic
qb_rol:MUSGAA             +  277.00  362.41  6.0e-12   1965  ! M74515 Mouse GA binding pro
qb_rol:AF256217           +  276.50  359.37  8.8e-12   2688  ! AF256217 Mus musculus ets 1
gb_htq2:AC055828          +  276.50  357.61  1.2e-11   177126  ! AC055828 Homo sapiens clo
qb_rol:AF256216           +  275.50  358.23  1.0e-11   2652  ! AF256216 Mus musculus ets 1
qb_ov:XLPLIG              +  275.00  353.62  1.8e-11   4404  ! X66979 X.laevis mRNA XEELI.
qb_ov:GGENAFLI            +  274.00  354.04  1.8e-11   3468  ! Y14774 Coturnix coturnix mR
qb_ov:GGELIG-NCO          +  274.00  353.92  1.8e-11   4546  ! Y14774 Coturnix coturnix mR
qb_rol:MMFLI1             +  272.50  352.45  1.1e-11   1729  ! X59421 Mouse FLI-1 mRNA for
qb_in3:DMELSO             +  272.50  354.16  2.0e-11   3006  ! X68259 D.melanogaster ETS-q
qb_ov:GGERGETS1A          +  272.00  352.60  1.1e-11   1531  ! M21823 Chicken c-ets 1 onco
qb_rol:MMTEL              +  270.50  349.29  3.2e-11   3512  ! Y07915 M.musculus mRNA for
qb_pr5:AF116510           +  269.50  354.84  1.6e-11   1415  ! AF116510 Homo sapiens Ets 1
qb_pr5:AF218465           +  268.50  354.58  1.9e-11   1400  ! AF218465 Homo sapiens Ets 1
qb_rol:MUSETS2            +  267.50  351.46  2.4e-11   1552  ! J04103 Mouse erythroblast s
qb_pr10:HUMFLI1A          +  267.50  349.80  3.0e-11   1932  ! M93255 Human FLI-1 mRNA, co
qb_ov:AF177538            +  266.50  351.79  2.3e-11   1245  ! AF177538 Danio rerio fli1 p
qb_pr5:AF175387           +  266.50  350.14  2.9e-11   1547  ! AF175387 Homo sapiens trans
qb_pat1:AR103484          +  266.00  339.63  1.1e-10   5667  ! AR103484 Sequence 4 from pa
qb_htq2:AC009371          +  266.00  316.67  2.1e-09   117175  ! AC009371 Drosophila melan
qb_pr10:HSU73844          +  265.50  347.46  4.1e-11   1846  ! U73844 Homo sapiens epithel
qb_pr5:AF124438           +  265.50  345.11  5.5e-11   2517  ! AF124438 Homo sapiens trans
qb_ov:DRU84615            +  265.50  343.23  9.0e-11   4260  ! U84615 Danio rerio E74 like
qb_ov:GGETSC              +  264.50  346.08  4.9e-11   1856  ! X14027 Chicken mRNA for c-e
qb_rol:BC005486           +  264.50  341.28  9.0e-11   4496  ! BC005486 Mus musculus, E2F
qb_rol:BC005504           +  264.00  340.75  9.6e-11   4434  ! BC005504 Mus musculus, E2F
qb_pat2:189256            +  263.00  339.85  1.1e-10   4240  ! 189256 Sequence 3 from pate
qb_pat1:AR103482          +  262.00  334.59  2.1e-10   5427  ! AR103482 Sequence 2 from pa
qb_pat1:AR103483          +  262.00  334.48  2.2e-10   5510  ! AR103483 Sequence 3 from pa
qb_in3:DROE74B            +  261.00  334.41  2.5e-10   5315  ! M87083 Drosophila melanogas
qb_in3:DROE74A            +  261.00  333.91  3.0e-10   6479  ! M87082 Drosophila melanogas
qb_ov:GGCETS              +  260.50  341.11  9.2e-11   1742  ! X14026 Chicken c-ets mRNA f
qb_ov:XLCETS2P            +  260.00  337.96  1.4e-10   2476  ! X52645 X.laevis mRNA for XL
qb_pr5:AF256222           +  260.00  337.34  1.5e-10   2653  ! AF256222 Homo sapiens ets 1
qb_pr10:HSU43188          +  260.00  335.86  1.8e-10   4222  ! U43188 Human Ets transcript
qb_pr10:HSU11732          +  259.50  340.60  9.8e-11   1580  ! U11732 Human ets-like gene
qb_pr5:AF212848           +  258.00  336.62  1.6e-10   2047  ! AF212848 Homo sapiens Ets d
qb_pr5:AF203977           +  258.00  333.33  2.5e-10   3160  ! AF203977 Homo sapiens ETS t
qb_rol:AF035527           +  256.50  332.94  1.4e-10   1319  ! AF035527 Mus musculus EBP (
qb_pr5:AF124439           +  256.00  332.17  2.9e-10   2586  ! AF124439 Homo sapiens trans
qb_pr5:AF170583           +  256.00  332.15  2.9e-10   2590  ! AF170583 Homo sapiens ets b
qb_in3:DM74E              -  256.00  327.48  5.3e-10   4800  ! X15087 Drosophila melanogas
qb_htq6:AC016512          -  255.50  308.83  5.8e-09   51405  ! AC016512 Drosophila melano
gb_in1:AE003523           -  255.50  295.89  3.0e-08   283337  ! AE003523 Drosophila melan
qb_pr10:HSU73843          +  255.00  333.10  2.6e-10   1915  ! U73843 Homo sapiens epithel
qb_pr10:HSU97156          +  255.00  333.00  2.6e-10   1942  ! U97156 Homo sapiens epithel
qb_pr7:BC003569           +  255.00  332.84  2.7e-10   1983  ! BC003569 Homo sapiens, E74
qb_ov:GGKETS1             +  254.50  329.65  4.8e-10   2764  ! M22462 Chicken protein p54
qb_in3:CEF24A1            -  252.50  306.85  7.4e-09   49248  ! AF024491 Caenorhabditis el
qb_pat2:196206            +  251.50  329.75  3.9e-10   1604  ! 196206 Sequence 43 from pat
qb_pr9:HSCETS1            +  251.50  329.75  3.9e-10   1604  ! X14798 Human DNA for c-ets
qb_pr5:AF116508           +  251.00  329.14  4.3e-10   1592  ! AF116508 Homo sapiens Ets 1
qb_rol:MMETS1             +  250.50  327.12  5.5e-10   1902  ! X54954 M.musculus ets 1 mRN
qb_ov:XLCETS1A            +  250.50  321.42  1.1e-09   1042  ! X52692 Xen pus laevis c-ets
qb_pat1:AR020776          +  250.00  326.48  6.1e-10   1920  ! AR020776 Sequence 1 from pa
qb_pr4:AF016295           +  250.00  326.48  6.1e-10   1929  ! AF016295 Homo sapiens Ets 1
qb_pr4:AF017307           +  248.50  322.28  1.0e-09   2529  ! AF017307 Homo sapiens Ets 1
qb_pat1:AX049151          +  246.00  321.06  1.2e-09   1907  ! AX049151 Sequence 4 from Pa
qb_pr10:HSU66894          +  246.00  321.06  1.2e-09   1907  ! U66894 Human epithelium tes
qb_om:AF057716            +  245.00  318.05  1.8e-09   2478  ! AF057716 ovis aries transcr
qb_pr4:AF020670           +  245.00  314.09  2.9e-09   4097  ! AF020670 Homo sapiens elf 1
gb_pr10:HSU32645          +  245.00  313.25  3.1e-09   4190  ! U32645 Human myeloid elf 1
qb_ov:GGCETS2             +  244.50  318.96  1.6e-09   1930  ! X07202 Chicken mRNA for c-e
qb_pr10:HUMETS1A          +  243.50  319.79  1.4e-09   1450  ! J04101 Human erythroblastos
gb_pr4:AF078104           +  243.00  317.83  1.8e-09   1719  ! AF078104 Homo sapiens fibGA
qb_ov:XELETS2AF           +  243.00  317.22  2.0e-09   1861  ! M81683 Xenopus ets 2a proto
qb_rol:MUSETS1Q           +  242.50  318.41  1.7e-09   1475  ! M58482 Mouse avian erythrob
```

```
qb_rol:AF016714      +  242.50  811.58  4.1e-09   4586  ! AF016714 Mus musculus myeloid
qb_pr5:AF256223      +  242.00  813.94  3.0e-09   2404  ! AF256223 Homo sapiens ets fami
qb_rol:AF256218      +  242.00  813.94  3.0e-09   2406  ! AF256218 Mus musculus ets fami
qb_rol:AF256219      +  242.00  813.82  3.0e-09   2142  ! AF256219 Mus musculus ets fami
qb_pat2:I89255       +  242.00  812.33  4.7e-09   2975  ! I89255 Sequence 1 from patent
qb_pr10:HSU44189     +  241.00  810.94  4.4e-09   2993  ! U44189 Human Ets transcription
qb_ro2:RATETSONCA    +  241.00  807.06  7.2e-09   4991  ! L20981 Rat proto-oncogene (Ets
qb_rol:MMELK1GEN     +  240.50  812.43  4.7e-09   2286  ! X87257 M.musculus mRNA for elK
qb_pat1:AR104814     +  239.00  817.82  1.8e-09    848  ! AR104814 Sequence 838 from pate
qb_ov:AF047008       +  239.00  805.50  8.9e-09   4307  ! AF047008 Carassius auratus tra
qb_rol:AF016294      +  237.00  808.98  5.7e-09   1910  ! AF016294 Mus musculus Ets tran
em_pat:E09411        +  236.00  808.09  6.3e-09   1800  ! E09411 cDNA of ETS transcripti
qb_rol:AY004174      +  236.00  807.40  7.0e-09   1998  ! AY004174 Mus musculus ets-rela
qb_pr9:HSETSRP       +  236.00  806.81  7.5e-09   2130  ! X76184 H.sapiens mRNA for ets
qb_pr10:HSU35622     +  234.00  804.21  1.0e-08   2107  ! U35622 Homo sapiens EWS protei
qb_rol:AB030217      +  233.50  805.75  8.6e-09   1575  ! AB030217 Rattus norvegicus mRN
qb_rol:AB030216      +  233.50  804.99  9.4e-09   1740  ! AB030216 Rattus norvegicus mRN
qb_rol:AB030215      +  233.50  801.54  1.5e-08   2744  ! AB030215 Rattus norvegicus mRN
qb_pr10:HUMC1S       +  233.50  799.56  1.9e-08   3564  ! M82882 Human cis-acting sequen
qb_rol:MM019617      +  233.50  798.92  2.1e-08   4876  ! U19617 Mus musculus Ets family
qb_rol:BC005520      +  231.50  796.50  2.8e-08   4746  ! BC005520 Mus musculus, Similar
qb_pat1:AR105459     +  230.50  799.67  1.9e-08   2064  ! AR105459 Sequence 1 from paten
qb_pr10:HUMEIAF      +  230.50  799.67  1.9e-08   2064  ! D12765 Human mRNA for EIA-F, 2
qb_pat2:E06025       +  230.50  799.64  1.9e-08   2073  ! E06025 cDNA encoding a protein
qb_pr10:HSU18018     +  230.50  798.74  2.1e-08   2333  ! U18018 Human EIA enhancer bind
qb_ov:AF168007       +  230.50  797.42  2.5e-08   2779  ! AF168007 Danio rerio Ets relat
qb_ov:DRU84616       +  230.50  795.69  3.1e-08   3490  ! U84616 Danio rerio E74-like fa
qb_rol:MMPEA3        +  229.50  797.16  2.6e-08   2410  ! X64799 M.musculus PEA3 mRNA.
qb_ov:DRAJ3209       +  229.00  799.06  2.0e-08   1716  ! AJ003209 Danio rerio mRNA for
qb_pr1:AB016193      +  228.50  794.97  3.4e-08   2695  ! AB016193 Homo sapiens ELK1 mRN
qb_htq22:AL512329        227.50  262.04  2.3e-06 173888  ! AL512329 Homo sapiens chrom
qb_ov:AY029410       +  226.50  798.06  2.3e-08   1258  ! AY029410 Xenopus laevis ternar
qb_ov:AF168008       +  225.50  793.95  3.9e-08   1813  ! AF168008 Danio rerio PEA3 (po
qb_pr4:AF000672      +  225.00  800.05  1.8e-08    742  ! AF000672 Homo sapiens ELK vari
qb_ov:AF075705       +  225.00  794.30  3.7e-08   1584  ! AF075705 Gallus gallus ets dom
qb_pat1:AR073028     +  225.00  791.59  5.3e-08   2266  ! AR073028 Sequence 1 from patc

seq_name: qb_pat1:AX048148

seq_documentation_block:
LOCUS       AX048148      1894 bp    DNA              PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent W00070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets domain proteins in cancer
  JOURNAL   Patent: WO 0070092 A 1 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416..1423
                     /note="Human JKETS."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MGSAGPSTLCSVGTSHLLLPTSTVSPTSLEKAAAGAVGLESFFWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLIPMGKAFQELAGKELCAMSEEQFRQKSPLGPVLHAHLQTWKSAAWMAER
                     TSPCATHYCASISEFSWLTSPVDSSSQQDTRWYSFELLIEPHSYCPEPWLNASS
                     GIFKIEDSAQVAPTWCIPFNPPAMNYDELSPSTPQYYEPCTIPPFPPISQKLVYQFVHP
                     I"
BASE COUNT       468 a    653 c    571 g    302 t
ORIGIN
```

```
alignment scores:
            Quality: 1785.00              Length:     435
              Ratio:  5.328                 Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2 x AX048148   ..

 Align seq 1/1  to: AX048148  from: 1  to: 1894

       1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     416 ATGGGCAGCGGCCAGCCCGGGCTCTGAGCAGCGTATCCCCCAGCCACCTCCT 465

      17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     466 GCTGCCCCCCGACACGGTGTCGCGGACAGGCTTGGAGAAGGCGGCAGCGG 515

      34 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     516 GGGCAGTGGGTCTCGAGAGACGGGACTGGAGTCCCAGTCCACCCGCCACG 565

      51 ProGluGluGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     566 CCCGAGCAGGGCCTGTCGCCCTTCTACCTCTCCTACTTTGACATGCTGTA 615

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     616 CCCTCAGCACACGCACTGGGCAGCCAAGCCCCCTGGGCCAGCACTCGGG 665

      84 luGlaProProGluGluProGluGlnCysProValIleAspSerGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     666 ACGAGCCACCTGACGCACCCTCACCAGTGCCCGGTCATTGACAGCCAAGCC 715

     101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHi 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     716 CCAGCGGGCCAGCCTGGACTTGGTGCCCGGCGGGCTGACCTTGGAGGAGCA 765

     117 sSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspI 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     766 CTCGCTGGAGCAGGTGCAGTCCATGGTGGTGGGCGAAGTGCTCAAGGACA 815

     134 leGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrp 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     816 TCGAGACGGCCTGCAAGCTGCTCAACATCACGGCAGATCCCATGGACTGG 865

     151 SerProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrAr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     866 AGCCCCAGCAATGTGGAGAAGTGGCTCCTGTGGACAGAGCACCAATACCG 915

     167 qLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuC 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     916 GCTGCCCCCCATGGGCAAGGCCTTCCAGGAGCTGGCGGGCAAGGAGCTGT 965

     184 ysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     966 GCGCCATGTCGGAGGAGCAGTTCCGCCAGCGCTCGCCCCTGGGTGGGGAT 1015

     201 ValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1016 GTGCTGCACGCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAAAGA 1065

     217 nArgIleSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluS 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1066 GCGGACTTCACCTGGGGGGATTCACTACTGTGCCTCGACCAGTGAGGAGA 1115

     234 erIleThrAspSerGlnValAspSerSerCysSerGlyGlnIleIleHis 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1116 GCTGGACCGACAGCGAGGTGGACTCATCATGCTCCGGGCAGCCCATCCAC 1165

     251 LeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyAr 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    1166 CTGTGGCAGTTCCTCAAGGAGTTGCTACTCAAGCCCCACAGCTATGGCCG 1215

     267 qPheIleArqTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspS 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1216 CTTCATTAGGTGGCTCAACAAGGACAAGGGCATCTTCAAAATTGAGGACT 1265

     284 erAlaGlnValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMet 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1266 CAGCCCAGGTGGCCCGGCTGTGGGGCATCCGCAAGAACCGTCCCGCCATG 1315

     301 AsnTyrAspLysLeuSerArqSerIleArqTyrTyrTyrLysLysGlyIl 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1316 AACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAAGGGCAT 1365

     317 eIleArqLysProAspIleSerGlnArqLeuValTyrGlnPheValHisP 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1366 CATCCGCAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACC 1415

     334 roIle 335
         |||||
    1416 CCATC 1420

seq_name: qb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538     1894 bp    mRNA            PRI       20-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
            Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium-specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate-specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1..1894
                     /gene="PDEF"
     CDS             416..1423
                     /gene="PDEF"
                     /note="prostate epithelium-specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLPSVSPSHLTRTDVSAPDPSLAKAVSEETSRFEAM
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGPVLKDIFTACKLINTDADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKFLCAMSFFQFRQRSPLQGDVIHARLDIWKSAAWMEER
                     TSPGATHYCASISLPFSWITGLVDSSCSQLFLNLWQFLLTLDRHSIGEFLFWLNEEM
                     GIFKTEDSAQVAPLWGIPKNPPAMNYDFLSRSIPQYYEACLIPEEFDISQRLVYQFVHP
                     I"
BASE COUNT       468 a    653 c    571 g    302 t
```

```
ORIGIN


alignment_scores:
            Quality: 1785.00              Length:     335
              Ratio:  5.328                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2 x AF071538   ..

 Align seq 1/1  to: AF071538  from: 1  to: 1894

       1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     416 ATGGGCAGCGCCAGCCCGGGTCTGAGCAGCGTATCCCCGAGCCACCTCCT 465

      17 uLeuProProAspThrValSerArqThrGlyLeuGluLysAlaAlaAlaG 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     466 GCTGCCCCCCGACACGGTGTCGCGGACAGGCTTGGAGAAGGCGGCAGCGG 515

      34 lyAlaValGlyLeuGluArqArqAspTrpSerProSerProProAlaIhr 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     516 GGGCAGTGGGTCTCGAGAGACGGGACTGGAGTCCCAGTCCACCCGCCACG 565

      51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     566 CCCGAGCAGGGCCTGTCCGCCTTCTACCTCTCCTACTTTGACATGCTGTA 615

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArqG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     616 CCCTGAGGACAGCAGCTGGGCAGCCAAGGCCCCTGGGGCCAGCAGTCGGG 665

      84 luGluProProGluGluProGluGlnCysProValIleAspSerGlnAla 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     666 AGGAGCCACCTGAGGAGCCTGAGCAGTGCCCGGTCATTGACAGCCAAGCC 715

     101 ProAlaGlySerLeuAspLeuValPr-GlyGlyLeuThrLeuGluGluHi 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     716 CCAGCGGGCCAGCCTGGACTTGGTGCCCGGCGGGCTGACCTTGGAGGAGCA 765

     117 sSerLeuGlnGlnValGlnSerMetValValGlyGluValLeuLysAspI 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     766 CTCGCTGGAGCAGGTGCAGTCCATGGTGGTGGGCGAAGTGCTCAAGGACA 815

     134 leGlnThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrp 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     816 TCGAGACGGCCTGCAAGCTGCTCAACATCACCGCAGATCCCATGGACTGG 865

     151 SerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrAr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     866 AGCCCCAGCAATGTGCAGAAGTGGCTCCTGTGGACAGAGCACCAATACCG 915

     167 qLeuProProMetGlyLysAlaPheGluGluLeuAlaGlyLysGluLeuC 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     916 GTTGCCCCCCATGGGCAAGGCCTTCGAGGAGCTGGCGGGCAAGGAGCTGT 965

     184 ysAlaMetSerGluGluGlnPheArqGlnArqSerProLeuGlyGlyAsp 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     966 GCGCCATGTCGGAGGAGCAGTTCCGCCAGCGCTCGCCCCTGGGTGGGGAT 1015

     201 ValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1016 GTGCTGCACGCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAAAGA 1065

     217 uAsqProSerProGlyAlaIleTyrLysAlaSerThrSerGluGluS 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1066 GCGGACTTCAGCCTGGGGCGATTCACTACTGTGCCTCCACCAGTGAGGAGA 1115

     234 erTrpThrAspSerGlyValAspSerSerCysSerGlyGluProIleHis 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1116 GCTGGACCGACAGCGAGGTGGACTCATCATGCTCCGGGCAGCCCATCCAC 1165
```

```
      253 LeuTrpGlnPheLeuLysGluGlyLeuLeuLysProIleSerTyrGlyAr 292
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1166 CTGTGGCAGTTCCTCAAGGAGTTGCTACTCAAGCCCCACAGCTATGGCCG 1215

      267 qPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGlnAspS 284
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1216 CTTCATTAGGTGGCTCAACAAGGAGAAGGGCATCTTCAAAATTGAGGACT 1265

      284 erAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMet 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1266 CAGCCCAGGTGGCCCGGCTGTGGGCATCCGCAAGAACCGTCCCGCCATG 1315

      301 AsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIl 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1316 AACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAAGGGCAT 1365

      317 eIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheValHisP 334
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1366 CATCCGGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACC 1415

      334 roIle 335
          |||||
     1416 CCATC 1420

seq_name: qb_pri:AB031549

seq_documentation_block:
LOCUS       AB031549     1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell line:PC 3 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20147504
REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1999) to the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases, 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pUC18"
     gene            406..1414
                     /gene="PSE"
     CDS             406..1414
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MetSerProLeuSerSerValSerProSerHisLeuLe
                     uProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAs
                     pSerProAlaThrProGlnGlnSerGlnGlnGlnSerProLeuGlyGlyAs
                     pGluLeuValProGlnLeuLeuProProSerGlnGlnSerMetValGlyGl
                     yHisLeuThrGlnAlaLeuGlnIleArgGlyProProAsnAspProValTh
                     rAlaSerGlnLeuProSerProAsnHisLeuMetIleAspProHisThrTy
                     rSerAspMetProAlaThrSerIleLeuGlnTrpThrValLeuProGluGl
                     yGlyProProAsnThrLeuIleHisTyrCysAlaSerThrSerGlnGluS
                     erProProProGlyThrGlnTyrIleLeuSerProLeuSerProValProA
                     snSerGlnAsnProArgSerGlnLysValThrLeuGluGlyAspMetGlyT
                     hrGlnLeuGlnLeuMetCysAlaGluArgGlyIleArgProGlnProGlnT
                     rpLeuTrpGlnPheLeuLysGluGlyLeuLeuLysProIleSerTyrGlyA
                     rgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGlnAspS
                     erAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetA
                     snTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleI
                     leArgLysProAspIleSerGlnArgLeuValTyrGlnPheValHisPro
                     IleProAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgProAla
                     MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGly
                     IleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheValHis
                     ProIleProAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgPro
                     AlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLys
                     GlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal
                     HisProIle"
BASE COUNT      380 a    651 c    567 g    300 t
ORIGIN


alignment_scores:
            Quality: 1785.00              Length:    835
              Ratio:  5.428                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2 x AB031549   ..

 Align seq 1/1  to: AB031549  from: 1  to: 1898

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      406 ATGGGCAGCGCCAGCCCGGGTCTGAGCAGCGTATCCCCCAGCCACCTCCT 455

       17 qLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      456 GCTGCCCCCGACACGGTGTCGCGGACAGGCTTGGAGAAGGCGGCAGCGGG 505

       34 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      506 GGGCAGTGGGTCTCGAGAGACGGGACTGGAGTCCCAGTCCACCCGCCACG 555

       51 ProGlnGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuly 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      556 CCCGAGCAGGGCCTGTCCGCCTTCTACCTCTCCTACTTTGACATGCTGTA 605

       67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      606 CCCTGAGGACAGCAGCTGGGCAGCCAAGGCCCCTGGGGCCAGCAGTCGGG 655

       84 luGluProProGluGluProGluGlnCysProValIleAspSerGluAla 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      656 AGGAGCCACCTGAGGAGCCTGAGCAGTGCCCGGTCATTGACAGCCAAGCC 705

      101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHi 117
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      706 CCAGCGGGCAGCCTGGACTTGGTGCCCGGCGGGCTGACCTTGGAGGAGCA 755

      117 sSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspI 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      756 CTCCCTGCAGCACCTCCACTCCATGGTGGTGGGCGAAGTGCTCAAGGACA 805

      134 leGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrp 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      806 TCGAGACGGCCTGCAAGCTGCTCAACATCACGGCAGATCCCATGGACTGG 855

      151 SerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrAr 167
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      856 AGCCCCAGCAATGTGCAGAAGTGGCTCCTGTGGACAGAGCACCAATACCG 905

      167 qLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuC 184
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      906 GCTGCCCCCCATGGGCAAGGCCTTCCAGGAGCTGGCGGGCAAGGAGCTGT 955

      184 ysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      956 GCGCCATGTCGGAGGAGCAGTTCCGCCAGCGCTCGCCCCTGGGTGGGGAT 1005

      201 ValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGl 217
          ||||||||||||||||||||||||||||||||||||||||||||||||||
     1006 GTGCTGCACGCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAAAGA 1055

      217 nArgProSerProGlyAlaIleHisTyrCysAlaSerThrSerGlnGluS 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    1056 GCGGACTTCACCTGGGGCGATTCACTACTGTGCCTCGACCAGTGAGGAGA 1105

     244 erIrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHis 250
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1106 GCTGGACCGACAGCGAGGTGGACTCATCATGCTCCGGGCAGCCCATCCAC 1155

     251 LeuIrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerIyrGlyAr 267
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1156 CTGTGGCAGTTCCTCAAGGAGTTGCTACTCAAGCCCCACAGCTATGGCCG 1205

     267 qPheIleArqIrpLeuAsnLysGlutysGlyIlePheLysIleGluAspS 284
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1206 CTTCATTAGGTGGCTCAACAAGGAGAAGGGCATCTTCAAAATTGAGGACT 1255

     284 erAlaGlnValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMet 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1256 CAGCCCACGTGGCCCGCTGTGGGCATCCGCAAGAACCGTCCCGCCATG 1305

     301 AsnIyrAspLysLeuSerArqSerIleArqGlnIyrIyrLysLysGlyIl 317
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1306 AACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAAGGGCAT 1355

     317 elleArqLysProAspIleSerGlnArqLeuValTyrGlnPheValHisP 334
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1356 CATCCGGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACC 1405

     334 rolle 335
         |||||
    1406 CCATC 1410

Seq_name: db_pat1:AX048159

seq_documentation_block:
LOCUS       AX048159     3317 bp    DNA             rAt       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
REFERENCE   1  (bases 1 to 3317)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 12 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note "GSEF-encoding sequence with promoter (Figs. 2A 2B)"
     promoter        1. .1380
                     /note "Promoter"
     CDS             1815. .2822
                     /note "GSEF coding sequence"
                     /codon_start=1
                     /protein_id="CAC19004.1"
                     /db_xref="GI:11876976"
                     /translation="MGSASSGLSSVSPSHLLPPDTVSPTGLEKAAAGAVGLEPSGLWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEPPSLEQVQSMVVGFVLKDIETACFLNETMKMEWETSSNVQFKLL
                     WTEHQYRLPPMGKAFQFLAGKFLCAMSFEQVFRQRSPLGGPVLHAHLDIWKSAAWMKER
                     TSPGATHYCASTSFFSWTDSFVDSSCSGQDIRLWQFLFLLLFFRRSYGFETEWLNPEE
                     GIFKIEDSAQVARLWGIRKNKPAMNYDKLSRSIRQYYKKGIIKKPIDISQRLVYQFVHP
                     I"
BASE COUNT       710 a    1026 c     970 g     611 t
ORIGIN


alignment_scores:
            Quality: 1785.00             Length:     835
              Ratio: 5.328                 Gaps:       0
```

```
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2 x AX048159   ..

 Align seq 1/1  to: AX048159  from: 1  to: 3317

       1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1815 ATGGGCAGCGCCAGCCCGGGTCTGAGCAGCGTATCCCCAGCCACCTCCT 1864

      17 uLeuProProAspThrValSerArqThrGlyLeuGluLysAlaAlaAlaG 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1865 GCTGCCCCCCGACACGGTGTCGCGGGACAGGCTTGGAGAAGGCGGCAGCGG 1914

      34 lyAlaValGlyLeuGluArqArqAspTrpSerProSerProProAlaThr 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1915 GGGCAGTGGGTCTCGAGACGCGGACTGGAGTCCCAGTCCACCCGCCACG 1964

      51 ProGluGluGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1965 CCCGAGCAGGGCCTGTCCGCCTTCTACCTCTCCTACTTTGACATGCTGTA 2014

      67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArqG 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2015 CCCTGAGGACAGCAGCTGGGCAGCCAAGGCCCCTGGGGCCAGCAGTCGGG 2064

      84 luGluProProGluGluProGluGlnCysProValIleAspSerGlnAla 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2065 AGGAGCCACCTGAGGAGCCTGAGCAGTGCCCCGTCATTGACAGCCAAGCC 2114

     101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHi 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2115 CCAGCGGGCAGCCTGGACTTGGTGCCGGGAGGCCTGACCTTGGAGGAGCA 2164

     117 sSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspI 134
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2165 CTCCTTGGAGCACCTGCAGTCCATGGTTGTGGGCGAAGTGCTCAAGGACA 2214

     134 leGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrp 150
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2215 TCACACGGCCTGCAAGCTGCTCAACATCACGGCCACATCCCATGCACTGG 2264

     151 SerProSerAsnValGluLysIrpLeuLeuTrpThrGluHisGluTyrAr 167
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2265 ACCCCAGCAATGTGGAGAAGTGGCTCCTGTGGACACAGCACCAATACCG 2314

     167 qLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuC 184
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2315 GCTGCCCCCCATGGGCAAGGCCTTCCAGGAGCTGGCGGGCAAGGAGCTGT 2364

     184 ysAlaMetSerGluGluGlnPheArqGlnArqSerProLeuGlyGlyAsp 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2365 GCGCCATGTCGGAGGAGCAGTTCCGCCAGCGCTCGCCCCTGGGTGGAGAT 2414

     201 ValIleuHisAlaHisLeuAspIleIrpLysSerAlaAlaIrpMetLysGl 217
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2415 GTGCTGCACGCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAAAGA 2464

     217 qArqTheSerIrpGlyAlaIleHisTyrCysAlaSerIhrSerGluGluS 234
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2465 GCGCACTTCACCTGGGGCGATTCACTACTGTGCCTCGACCAGTGAGGAGA 2514

     234 erIrpIhrAspSerGluValAspSerSerCysSerGlyGlnProIleHis 250
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2515 GCTGGACCGACAGCGAGGTGGACTCATCATGCTCCGGGCAGCCCATCCAC 2564

     251 LeuIrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyAr 267
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2565 CTGTGGCAGTTCCTCAAGGAGTTGCTACTCAAGCCCCACAGCTATGGCCG 2614

     267 qPheIleArqIrpLeuAsnLysGluLysGlyIlePheLysIleGluAspS 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     2615 CTTCATTAGGTGGCTCAACAAGGAGAAGGGCATCTTCAAAATTGAGGACT 2664

      284 crAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMet  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     2665 CAGCCCAGGTGGCCCGGCTGTGGGGCATCCGCAAGAACCGTCCCGCCATG 2714

      301 AsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIl  317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     2715 AACTACGACAAGCTGAGCCGCTCCATCCGCCAGTATTACAAGAAGGGCAT 2764

      317 cIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheValHisP  334
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     2765 CATCCGAAGCCAGACATCTCCCAGCGCCTCGTCTACCAGTTCGTGCACCC 2814

      334 rolle 335
          |||||
     2815 CCATC 2819

seq name: qb rol:AB019436

seq documentation block:
locUS       AB019436      1704 bp     mRNA            ROD        20 JAN 2000
DEFINITION  Mus musculus mRNA for Prostate specific transcription factor ets,
            complete cds.
ACCESSION   AB019436
VERSION     AB019436.1  GI:4760536
KEYWORDS    Prostate specific transcription factor ets; mPse.
SOURCE      Mus musculus (strain:C57BL/6) Adult male colon Epithelial cDNA to
            mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqmathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267 274 (2000)
  MEDLINE   20147504
REFERENCE   2  (bases 1 to 1704)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02 NOV 1998) to the DEBAJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6 879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1. .1704
                     /organism "Mus musculus"
                     /strain "C57BL/6"
                     /db_xref "taxon:10090"
                     /cell_type "Epithelial"
                     /dev_stage "Adult"
                     /sex "male"
                     /tissue_type "Colon"
     gene            362. .1349
                     /gene "mPse"
     CDS             362. .1349
                     /gene "mPse"
                     /codon_start 1
                     /product "Prostate specific transcription factor ets"
                     /protein_id "BAA77329.1"
                     /db_xref "GI:4760537"
                     /translation "MGSASPGLSNVSPGCLLLFPDVAPKTGIEKAASGAMQPEKQEWS
                     PSPPATPEQGLSAFYLSYFNMYPDDSSWVAKVPEAKAGEDIPEEPEQCPVIDSQASGS
                     TLDEHSLEQVQSMVVGEVLKDIETACKLLNTTADPGDWSPGNVQKWLLWTEHQYRLPP
                     AGKAFQELGGKELCAMSEEQFRQKAPLGGDVDLHAHLDTWKSAAWMKERTSPGTLHYCA
                     STSFEGWTQGEVDSSCSSQPTHLWQFLKELLLKPHSYGPFIRWLNKEKGLELEDSAQ
                     VARLWGVRKNRPAMNYDKLSRSIRQYYKKGIIRKPDPTSQRLVYQFVHPV"
     polyA_site      1704
                     /note "10 a nucleotides"
BASE COUNT      369 a    527 c    491 g    317 t
ORIGIN
```

```
alignment scores:
            Quality: 1531.00              Length:     435
              Ratio:  4.891                 Gaps:       2
 Percent Similarity:  93.433   Percent Identity: 85.672

alignment block:
 US-09-126-945B-2 x AB019436   ..

 Align seq 1/1  to: AB019436  from: 1  to: 1704

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe  17
          ||||||||||||||||||||||||||||||:::|||||||||:::   |||||
      462 ATGGGCAGTGCCAGCCCAGGCCTGAGCAACGTGTCCCCCGGTTGCCTGCT 411

       17 uLeuProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaG  34
          ||||   |||||||||||:   |||||||||   |||||||||||||:||
      412 ACTGTTCCCAGATGTGGCACCACGAACAGGGACGGAGAAGGCAGCATCAG 461

       34 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaThr  50
          |||||:|||   |||:::::::::|||||||||||||||||||||||||
      462 GAGCAATGGGCCCTGACAAGCAGGAATGGAGTCCTAGTCCACCCGCCACC 511

       51 ProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy  67
          ||||||||||||||||||||||||||||||||||||||:::|||   ||
      512 CCTGAGCAGGGCCTGTCTGCTTTCTACCTCTCTTACTTTAACATG...TA 558

       67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG  84
          ||||:::|||||||||||||:::|||||:::|||   |||   :::   |
      559 TCCCGACGATAGCAGCTGGGTCGCCAAAGTCCCCGGAGGCCCGTGCCGGGG 608

       84 luGluProProGlnGluProGluGlnCysProValIleAspSerGlnAla 100
          ||||:   |||||||||||||||||||||||||||||||||||||||||||
      609 AGGACCACCCCGGAGGAGCCCGAGCAGTGTCCCGTCATTGACAGCCAGGCC 658

      101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHi 117
                                     :::|||   ||||||:::||||||
      659 .........................TCTGGGAGCACGTTGGATGAGCA 681

      117 sSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspI 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      682 CTCGCTAGAGCAGGTGCAATCGATGGTTGTGGGCCAGGTCCTGAAAGATA 731

      134 leGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrp 150
          ||||||||||||||||||||||||||||||||||||||||   ||||||
      732 TTGAGACGGCCTGCAAGCTTCTGAACATCACAGCAGACCCTGGGGACTGG 781

      151 SerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrAr 167
          ||||||:::|||||||||||||||||||||||||||||||||||||||||
      782 AGCCCTGGTAACGTGCAGAAGTGGCTTTTATGGACAGAACACCAGTACCG 831

      167 qLeuProProMetGlyLysAlaPheGlnLeuLeuAlaGlyLysGluLeuC 184
          ||||||||||   |||||||||||||||||||||:::|||||||||||||
      832 GCTGCCTCCAGCAGGCAAGGCCTTCCAGGAGCTGGGCGGTAAGGAGCTGT 881

      184 ysAlaMetSerGluGluGlnPheArgGluArgSerProLeuGlyGlyAsp 200
          ||||||||||||||||||||||||||||||:::|||||||||||||||||
      882 GCGCCATGTCCGAGGAACAGTTCCGTGAGCGTGCACCCTTGGGTGGGGAT 931

      201 ValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGl 217
          ||||||||||:|||||||||||||||||||||||||||||||||||||||
      932 GTACTGCATGCCCACCTGGACATCTGGAAGTCAGCGGCCTGGATGAAGGA 981

      217 uArgThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluS 234
          ||||||||||||:::::|||||||||||||||||||||||||||||||||:
      982 GAGGACCTCTCGCCTGGCACCCTTCACTACTGCGCCTCCACCAGCGAGGAGG 1031

      234 erTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHis 250
          ::|||||||||:::|||||||||||||||||||||||||||||||||||||
     1032 GCTGGACGGATGGTGAGTGCACTCGTCGTGCTCCGGGCAGCCCATTCAC 1081

      251 LeuTrpGlnPheLeuLysGluLeuLeuLysProHisSerTyrGlyAr 267
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1082 CTGTGGCAGTTCCTGAAAGAACTGCTGCTCAAGCCCCACAGCTATGGCCG 1131

      267 qPheIleArgIrpLeuAsnLysGluLysGlyIlePheLysIleGluAspS 284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     1132 CTTCATCCGCTGGCTCAACAAGGAGAAAGGCATCTTCAAAATTGAGGACT 1181

      284 erAlaGlnValAlaArgLeuIrpGlyIleArgLysAsnArgProAlaMet 300
          ||||||||||||||||||||||||||:::||||||||||||||||||||||
     1182 CAGCACAGGTGGCCCGACTGTGGGGTGTGCGCAAGAACCGGCCAGCCATG 1231

      301 AsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIl 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     1232 AACTATGATAAACTAAGCCGCTCCATCCGCCAGTATTACAAGAAGGGCAT 1281

      317 eIleArgLysProAspIleSerGlnArgLeuValIyrGluPheValHisP 334
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
     1282 CATTCGTAAACCCGACATCTCTCAGCGCCTTGTCTACCAATTTGTCCATC 1331

      334 roIle 335
          |||::
     1332 CAGTC 1336

seq_name: qb_pr6:AL157372

seq_documentation_block:
LOCUS       AL157372   150660 bp    DNA              PRI       29-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-375E1 on chromosome 6, complete
            sequence.
ACCESSION   AL157372
VERSION     AL157372.18  GI:11493236
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150660)
  AUTHORS   Skuce,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11342772.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-375E1 is from the library RPCI 11.2 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-375E1 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-375E1 is at 1 in this sequence. The
            true left end of clone RP3-391O22 is at 150561 in this sequence.
            The true right end of clone RP11-302L10 is at 113048 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1..150660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-375E1"
                     /clone_lib="RPCI-11 2"
BASE COUNT     35470 a    38887 c    38740 g    37563 t
ORIGIN


alignment_scores:
            Quality:  790.50              Length:   2061
              Ratio:  2.374                 Gaps:      7
 Percent Similarity:  16.157   Percent Identity: 16.109

alignment_block:
 US-09-126-945B-2 x AL157372/rev  ..

 Align seq 1/1  to reverse of: AL157372  from: 1  to: 150660

        1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
   129740 ATGGGCAGCGCCAGCCCGGGTCTGAGCAGCGTATCCCCAGCCACCTCCT 129691

       17 eLeuProProAspThrValSerArgIhrGlyLeuGlyLysAlaAlaAlaG 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
   129690 GCTGCCCCCCGACACCGTGTCGCGGACAGGCTTGGAGAAGCGGCGGCAGCGG 129641

       34 lyAlaValGlyLeuGluArgArgAspTrpSerProSerProProAlaIhr 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
   129640 GGGCAGTGGGTCTCGAGAGACGGGACTGGAGTCCCAGTCCACCCGCCACG 129591

       51 ProGluGluGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTy 67
          ||||||||||||||||||||||||||||:|||||||||||||||||||||||
   129590 CCCGAGCAGGGGCTGTCGGCTTCTACCTCTCCTACTTTGACATGCTGTA 129541

       67 rProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
   129540 CCCTCAGGACAGCAGCTGGGCAGCCAAGGCGCCTGGGGCCAGCAGCTCGGC 129491

       84 luGluProProGluGluProGluGlnCysProValIleAspSerGlnAla 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
   129490 AGGAGCCAGCCTGAGGAGCCTGAGCAGTGCCCGGTCATTGACAGCCAAGCC 129441

      101 ProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluTi 117
          |||||:||||||||||||||||||||||||||||||||||||||||||||||
   129440 CCAGCGGGCAGCCTGGACTTGGTGCCCGGCGGGCTGACCTTGGAGGAGCA 129391

      117 sSerLeuGluGlnValGluSerMetValValGlyGluValLeuLysAspI 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
   129390 CTCGCTGGAGCAGGTGCAGTCCATGGTGGTGGGCGAAGTGCTCAAGGACA 129341

      134 leGluThrAlaCysLysLeuLeuAsnIleThrAla..............  145
          ||||||||||||:||||||||||||||||||||
   129340 TCGACACCGCCTGTAAGCTCCTCAACATCACCGCACGTGACCCCTCTCCA 129291

      145 ..................................................  145

   129290 TGTGCCTGCTTCGCTAAGATGGGTCACAGGCAAGGCTGAGGTCGCTGGGG 129241

      145 ..................................................  145

   129240 TGGGTCGGGGCACTGCTGGATGGACTCTGGCCACCTCTGGCCCAGCCTCG 129191

      145 ..................................................  145

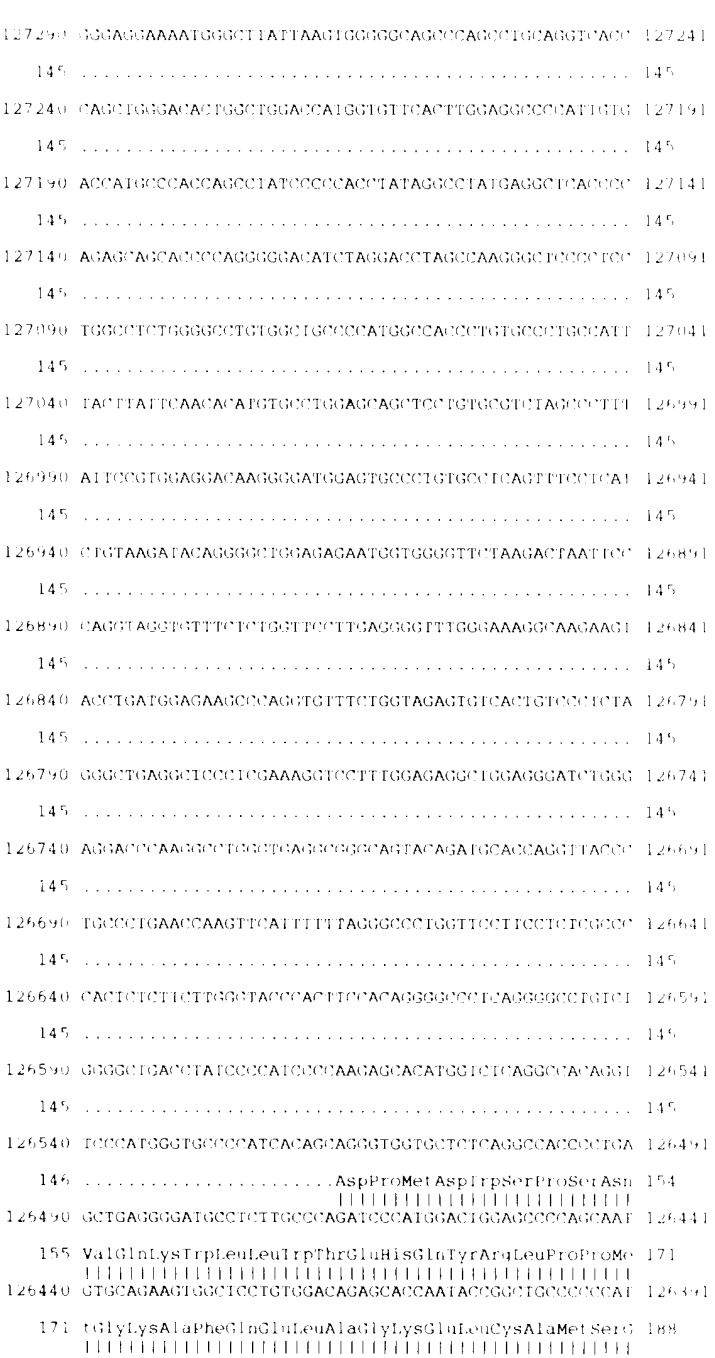
   129190 GGCAGCTTTGGGGATGCTGGCACTTCTACACACAGTCAGCTGTGAGCCTG 129141

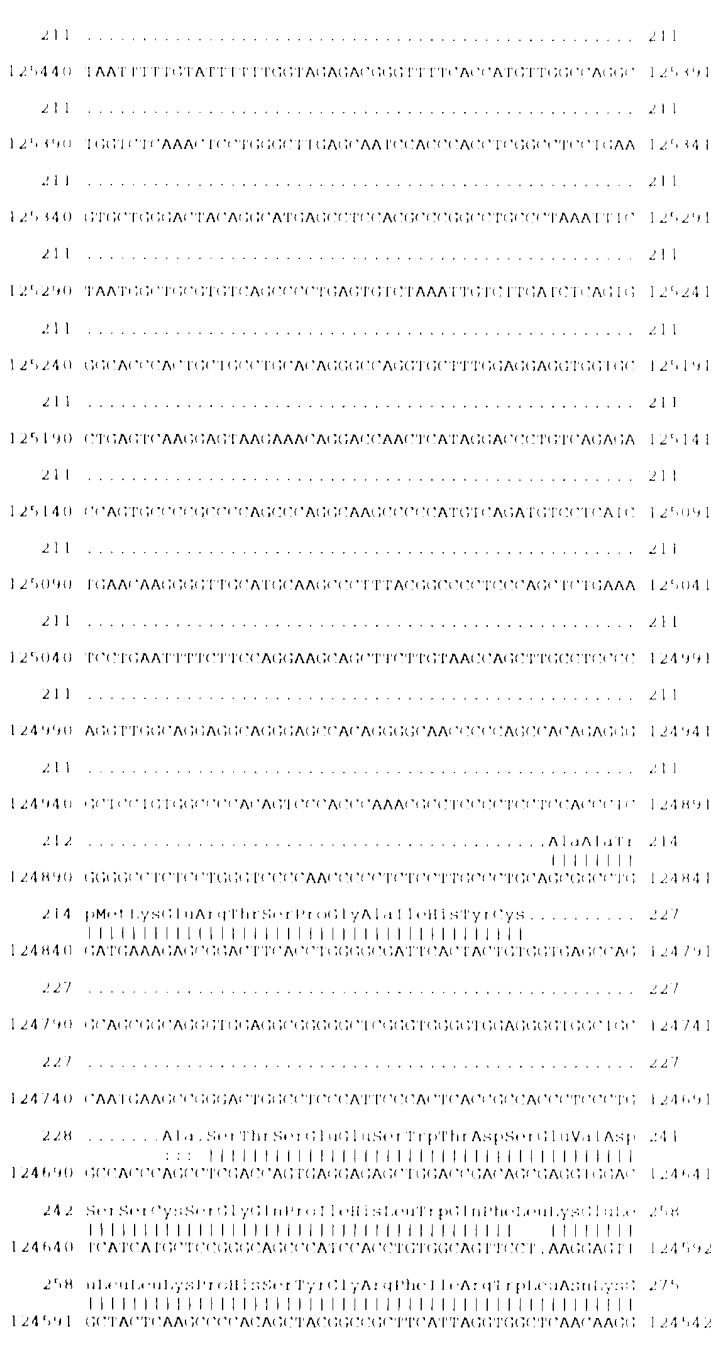
      145 ..................................................  145

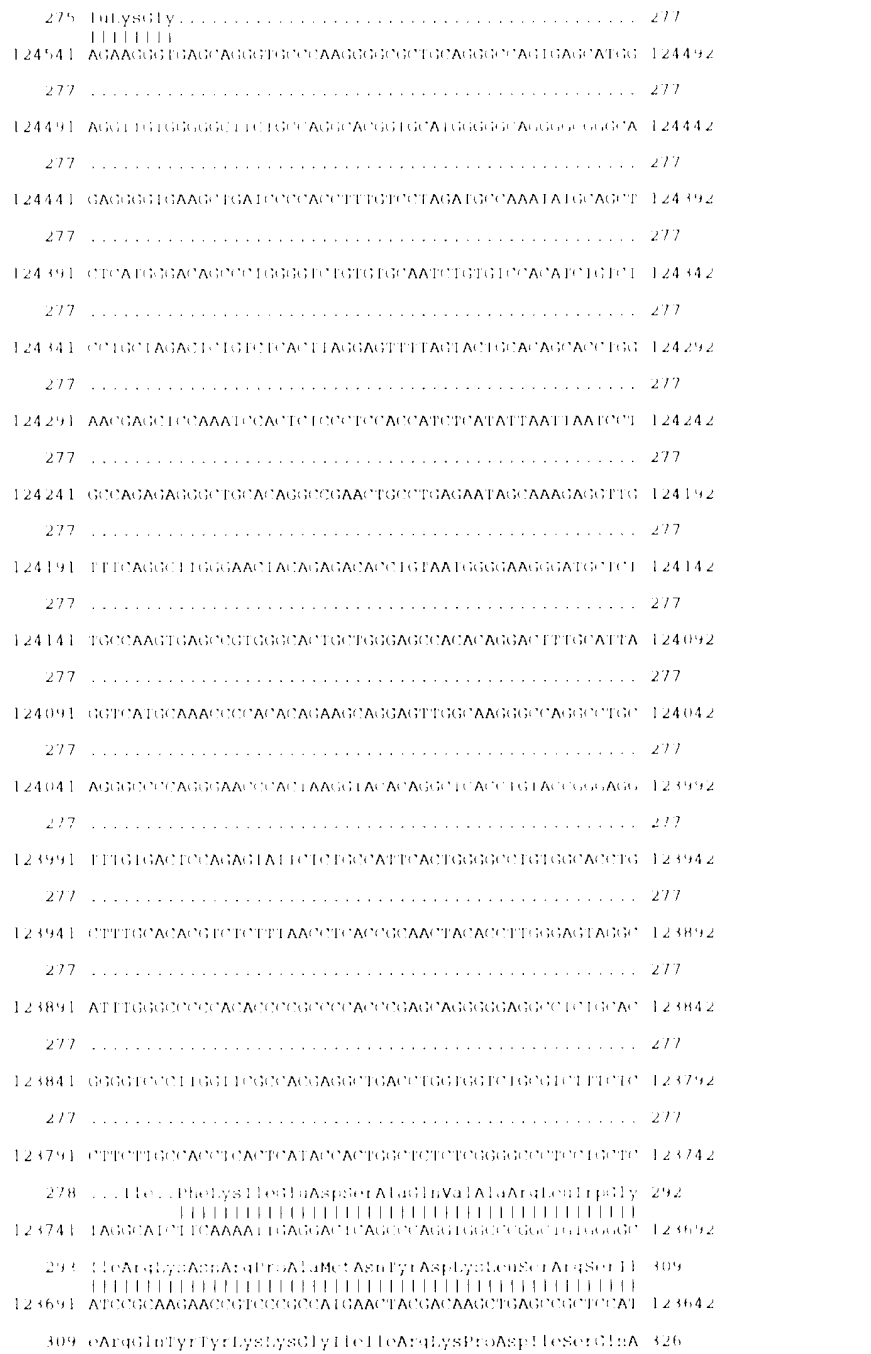
   129140 GGCCAGGAAGGTTTCTGTCCACCTCTAGTGTCCCAACCATGCCAGCCTAG 129091
```

```
   145 ..................................................  145
129090 GGTCCCCACTGGTAGCTGCTGGATGATATGAATTCAAGCTCTGTGTCCCC  129041
   145 ..................................................  145
129040 CACCTCCTCTTGGACTTCGCTCTCCCCTCATTCACTCTGCTCCCGCCACA  128991
   145 ..................................................  145
128990 CTGGCCTGTTTGCTGTTCCCACAACATTTGCTACGCTGCTGCTTCTTCAG  128941
   145 ..................................................  145
128940 CCTAGAGCGCTCTTCCCCCAAATATCTGTGTGACTCACTTCCTTCAGGTC  128891
   145 ..................................................  145
128890 TTGGCTCAGATGGCACCATCCGCATGAGGCCTTCCCTGACCCCTCCATTA  128841
   145 ..................................................  145
128840 AAATTGCTGCCCATGAGCACAGGACATGGTAAAGAAAACAAACAAAAAAA  128791
   145 ..................................................  145
128790 TGAAAAATAAAAGAAAGAAGAAAAATTTCTGCTCACATCCCTACCTGCTG  128741
   145 ..................................................  145
128740 TTCTCACCTGCTTCCTCTCATAGCACTTACACTCCTCGGACGTGCCGCAC  128691
   145 ..................................................  145
128690 CATATCTGTTGATCTCGCTCACTCTCTGTGTCCTCCACTAGAATGTACAC  128641
   145 ..................................................  145
128640 TCCACAAAGGCAGGGATTTTCATTTGTTTTGCTCATCAATCCATCCCCAG  128591
   145 ..................................................  145
128590 ATGCTAGGATAGCACCTTGAATATAACTGATGCTCAATAGACAGTTGTTC  128541
   145 ..................................................  145
128540 AATAATAAGCATTTCTTTTTTTTTTCTTTTTCTTCTTTTTTTTTTTTTT  128491
   145 ..................................................  145
128490 TTTTTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGGCTGCAGTGCAGT  128441
   145 ..................................................  145
128440 GGTGCAATCTCGGCTCACTCACTGCAAGCTCTGCCTCCCGGGTTCAAGCC  128391
   145 ..................................................  145
128390 ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCCGCCAC  128341
   145 ..................................................  145
128340 CACACCCGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCGTGT  128291
   145 ..................................................  145
128290 TAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCTGCCTGCCTCGGCC  128241
   145 ..................................................  145
128240 TCCCAAAGTGCTGGGATTACTAGCGTGAGCCACCGCGCCCAGCCTCTTCT  128191
```

```
   145 ..................................................  145
128190 TTTTTTTTTTTTTTTTTTGACAGGGTCTCGCTCTATCACTGAGGCTGGA  128141
   145 ..................................................  145
128140 GTGCAGTGGTGTGATTTCAGCTCACTGCAACCTCTGCCTCTGGGTTCAAG  128091
   145 ..................................................  145
128090 CGATTCTCCTGCCTCAGCCTCCTGAGAAGCTGGGATTACAGGTGTCTGCC  128041
   145 ..................................................  145
128040 ACCACACCCTGCTAATTTTTATATTTTTAGTAGAGACGGAGTTTTCACCA  127991
   145 ..................................................  145
127990 TGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTCATCCGCCTGCCT  127941
   145 ..................................................  145
127940 TGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCTCTCTGGCCCC  127891
   145 ..................................................  145
127890 AAAAGCAGTTCCTAATATGCCAAGCCTTATTCTAGGACGGGGGAATTGAG  127841
   145 ..................................................  145
127840 TGGTGAACACAACACAGCCCCTGTGTTCGTAGGGCTGACATGATAGTGAG  127791
   145 ..................................................  145
127790 TGACACAGATCACAAATAAATACAGAAATGTGTCACTCTTCTTCTGGGAC  127741
   145 ..................................................  145
127740 GGAAGGCCCCGATGACGAGGGGACCTCCTAGGGCCTGGCATACATGTGTC  127691
   145 ..................................................  145
127690 TGTCAGCCAGACCTGGAAGGCGTGATTATGGAGCTGATGCTGTGGGTGGC  127641
   145 ..................................................  145
127640 GGGGCTGCTGAGTCTCTAGTCCTCTGTCCTGCTGTCGCCTCTCCAGCTTC  127591
   145 ..................................................  145
127590 AGGCAGTCCTGACCCATGTCCCTCATCTTCAGGCCCTGGCATTGCAGCAC  127541
   145 ..................................................  145
127540 TGAACTCTGTGCAGCTTTCCCTGGGAAGCAGGGGTCTTCATGCATCCCTC  127491
   145 ..................................................  145
127490 AGGCCCGGCCCCTGCCCCCAAAGCTGCCCCGTTGACGTCCTCATTGAAG  127441
   145 ..................................................  145
127440 CTGCTCCTGCCGATGCCCCAAGTTTCCCAATGCCCATGCCAGGCCGACTC  127391
   145 ..................................................  145
127390 CAGGAGTTCTCCCACAGACCTCCTCCCCACCCACAGGGCTAGGGAGGGGA  127341
   145 ..................................................  145
127340 GGGCCCTGCTCCCTCAACAATGGGAACAGAAACAAATGGCTTTGACCTCA  127291
   145 ..................................................  145
```


127290 GGGAGGAAAATGGGCTTATTAAGTGGGGGCAGCCCAGCCTGCAGGTCACC 127241
145 145
127240 CAGCTGGGACACTGGCTGGACCATGGTGTTCACTTGGAGGCCCCATTGTG 127191
145 145
127190 ACCATGCCCACCAGCCTATCCCCCACCTATAGGCCTATGAGGCTCACCCC 127141
145 145
127140 AGAGCAGCACCCCAGGGGGACATCTAGGACCTAGCCAAGGGCTCCCCTCC 127091
145 145
127090 TGGCCTCTGGGGCCTGTGGCTGCCCCATGGCCACCCTGTGCCCTGCCATT 127041
145 145
127040 TACTTATTCAACACATGTGCCTGGAGCAGCTCCTGTGCGTCTAGCCCTTT 126991
145 145
126990 ATTCCGTGGAGGACAAGGGATGGAGTGCCCTGTGCCTCAGTTTCCTCAT 126941
145 145
126940 CTGTAAGATACAGGGGCTGGAGAGAATGGTGGGGTTCTAAGACTAATTCC 126891
145 145
126890 CAGGTAGGTGTTTCTCTGGTTCCTTGAGGGGTTTGGGAAAGGCAAGAAGT 126841
145 145
126840 ACCTGATGGAGAAGCCCAGGTGTTTCTGGTAGAGTGTCACTGTCCCTCTA 126791
145 145
126790 GGGCTGAGGCTCCCTCGAAAGGTCCTTTGGAGAGGCTGGAGGGATCTGGG 126741
145 145
126740 AGGACCCAAGGCCTGGGTGAGGCGGCAGTACAGATGCACCAGGTTACCC 126691
145 145
126690 TGCCCTGAACCAAGTTCATTTTTTAGGGCCCTGGTTCCTTCCTCTCGCCC 126641
145 145
126640 CACTCTCTTCTTGGGTACCCACTTCCACAGGGGCCCTCAGGGGCCTGTCT 126591
145 145
126590 GGGGCTGACCTATCCCCATCCCCAAGAGCACATGGTCTCAGGCCACAGGT 126541
145 145
126540 TCCCATGGGTGCCCCATCACAGCAGGGTGGTGCTCTCAGGCCACCCCTGA 126491
146 ..... AspProMetAspTrpSerProSerAsn 154
||||||||||||||||||||||||||
126490 GCTGAGGGGATGCCTCTTGCCCAGATCCCATGGACTGGAGCCCCAGCAAT 126441
155 ValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMe 171
|||||||||||||||||||||||||||||||||||||||||||||||||||
126440 GTGCAGAAGTGGCTCCTGTGGACAGAGCACCAATACCGGCTGCCCCCCAT 126391
171 tGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerG 188
|||||||||||||||||||||||||||||||||||||||||||||||||||
126390 GGGCAAGGCCTTCCAGGAGCTGGCGGGCAAGGAGCTGTGCGCCATGTGG 126441
188 TuGluGlnPheArgGluArgSerProLeuGlyGlyAspValLeuHisAla 204
|||||||||||||||||||||||||||||||||||||||||||||||||||
126340 AGGAGCAGTTCCGCCAGCGCTCCCCCTGGGTGGGGATGTGCTGCACGCC 126291
205 HisLeuAsp..rpLys..Ser..... 211
|||||||||||||||||
126290 CACCTGGACATCTGGAAGTCAGGTAGGATGTGGCATGGTGGGAGGAGCC 126241
211 211
126240 CCTCCCAGGCTCCAAGCCAGGGACAGGATGCTGGCTAGGAGGTTGCCCTC 126191
211 211
126190 TGCTAGGGCAGCAGCCCCTTGGCCCCCAACTGGACTCCACTCTGAATCTC 126141
211 211
126140 AGCCCCAGTTTCAGTTGCCTCTGGATCCTGGTGCACTCTTCACCCAAC 126091
211 211
126090 CCACTCCCAGCCCTGGTCAAGGCCCAGAGCTTTGGGACAGAGCACCTGG 126041
211 211
126040 GAGGTTAAAAGGTAGAAGTCAGACTGTGAGGATGCAGGATGTCAGAGCTG 125991
211 211
125990 GGGACAACCGTGGATACTAGGAAGTCCTTTTCCTTTTAAAATTTTGTTT 125941
211 211
125940 GTTTTTCTTTGTTTTCTTTGTTTTTTTTTATTAAAGAAGTAATAAATGTG 125891
211 211
125890 CATGTTATAAAATAGACCACAGTGAAAAGTAGTGTACCATGAACAGTTTC 125841
211 211
125840 CCCCTCCCGCTCCCGACCCCCAGCTTCCAATCCCCTCCCTAGAGGGAACC 125791
211 211
125790 ACTGTTACAAGTTTCTTGAGTATCTTTCCAGGGAAAGTCTATGAATACTC 125741
211 211
125740 GAGCACATTTACATATATGCAGTTCTTATTCTTTTTTTTTTTTTGACAT 125691
211 211
125690 AAGTAGTAGCAGAGTATTACACTGCTCTGCACACCACAGTATGTCTTGGG 125641
211 211
125640 TATTGCCCCTGTCCGTGTCCACAGTGGCCCTGTTTTAATTTTAGTTTTTG 125591
211 211
125590 TTTTTTGAGACAGGGTCTCACTCTATTGCCCAGGCTGGAGTACAGTGGCA 125541
211 211
125540 CAGTCACAGCTCACTATAGCCTCGACTTCCCTGTTTCATTGGATCCTCCC 125491
211 211
125490 ACCCCAATCTCCTGAGTAGCTGAGTCTACAGGTGCACCACCACACCCTGC 125441


211 211
125440 TAATTTTGTATTTTTGGTAGAGACGGGTTTTCACCATGTTGGCCAGGC 125391
211 211
125390 TGGTCTCAAACTCCTGGGCTTGAGCAATCCACCCACCTCGGCCTCCCAAA 125341
211 211
125340 GTGCTGGGATTACAGGCATGAGCCTCCACGCCCGGCCTGCCCTAAATTTC 125291
211 211
125290 TAATGGCTGCGTGTCAGCCCCTGAGTGTCTAAATTGTCTTGATCTCAGTG 125241
211 211
125240 GGCACCCACTGCTGCCTGCACAGGCCCAGGTGCTTTGGAGGAGTGGTGC 125191
211 211
125190 CTGAGTCAAGGAGTAAGAAACAGGACCAACTCATAGGACCCTGTCAGAGA 125141
211 211
125140 CCAGTGCCCCGCCCCAGCCCAGGCAAGCCCCCATGTCAGATGTCCTCATC 125091
211 211
125090 TGAACAAGGGGTTGCATGCAAGCCCTTTACGGCCCCTCCCAGCTCTGAAA 125041
211 211
125040 TCCTGAATTTTCTTCCAGGAAGCAGCTTCTTGTAACCAGCTTGCCTCCCC 124991
211 211
124990 AGGTTGGCAGGAGGCAGGAGCCACAGGGGCAACCCCCAGCCACAGAGGC 124941
211 211
124940 GCTCCTGTGGCCCCACAGTCCCACCCAAACGCCTCCCCTCCTCCACCCTC 124891
212 .....AlaAlaTr 214
124890 GGGGCCTCTCCTGGGTCCCCAACCCCCTCTCCTTGCCCTGCAGCGGCCTG 124841
214 pMetLysGluArgThrSerProGlyAlaIleHisTyrCys..... 227
124840 GATGAAAGAGCGGACTTCACCTGGGGCGATTCACTACTGTGGTGAGCCAG 124791
227 227
124790 GCAGCGGCAGGGTGGAGGCGGGGGCTCGGGTGGGGTGGAGGGGTGGCTGC 124741
227 227
124740 CAATGAAGCCGGGACTGGCCTCCCATTCCCACTCACCGCCACCCTCCCTG 124691
228 .....AlaSerThrSerGluGluSerTrpThrAspSerGluValAsp 241
124690 GCCACCCAGCCTCGACCAGTGAGGAGAGCTGGACCGACAGCGAGTGGAC 124641
242 SerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLe 258
124640 TCATCATGCTCCGGGCAGCCCATCCACCTGTGGCAGTTCCTLAAGGAGTT 124592
258 uLeuLeuLysProHisSerTyrCysArgPheIleArgTrpLeuAsnLysS 275
124591 GCTACTCAAGCCCCACAGCTACGGCGCTTCATTAGGTGGCTCAACAAGG 124542


275 luLysGly..... 277
124541 AGAAGGGTGAGCAGGGTGCCCAAGGGGCGCTGCAGGGCCAGTGAGCATGG 124492
277 277
124491 AGGTTGTGGGGGCTTCTGCCAGGCACGGTGCATGGGGCCAGGGGCGGCA 124442
277 277
124441 GAGGGGTGAAGCTGATCCCCACCTTTGTCCTAGATGCCAAATATGCAGCT 124392
277 277
124391 CTCATGGGACAGCCCTGGGGTCTGTGTGCAATCTGTGTCCACATCTGTCT 124342
277 277
124341 CCTGCTAGACTCTGTCTCACTTAGGAGTTTTAGTACTGCACAGCACCTGG 124292
277 277
124291 AACGAGCTCCAAATCCACTCTCCCTCCACCATCTCATATTAATTAATCCT 124242
277 277
124241 GCCAGAGAGGGCTGCACAGGCCGAACTGCCTGAGAATAGCAAAGAGGTTG 124192
277 277
124191 TTTCAGGCTTGGGAACTACAGAGACACCTGTAATGGGGAAGGGATGCTCT 124142
277 277
124141 TGCCAAGTGAGCCGTGGGCACTGCTGGGAGCCACACAGGACTTTGCATTA 124092
277 277
124091 GGTCATGCAAACCCCACACAGAAGCAGGAGTTGGCAAGGGCCAGGCCTGC 124042
277 277
124041 AGGGCCCCAGGGAACCCACTAAGGTACACAGGCTCACCTGTACCGGAGG 123992
277 277
123991 TTTGTGACTCCAGAGTATTCTCTGCCATTCACTGGGGCCTGTGGCACCTG 123942
277 277
123941 CTTTGCACACGTCTCTTTAACCTCACCGCAACTACACCTTGGGAGTAGGC 123892
277 277
123891 ATTTGGGCCCCCACACCCCGCCCCACCCGAGCAGGGGAGGCCTCTGCAC 123842
277 277
123841 GGGGTCCCTTGGTTCGCCACGAGGCTGACCTGGTGGTCTGCGTCTTTCTC 123792
277 277
123791 CTTCTTGCCACCTCACTCATACCACTGGCTCTCTCGGGGCCCTCCTGCTC 123742
278 ...Ile..PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGly 292
123741 TAGGCATCTTCAAAATTGAGGACTCAGCCCAGGTGGCCCGCCTGTGGGGC 123692
293 IleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIl 309
123691 ATCCGCAAGAACCGTCCCGCCATGAACTACGACAAGCTGAGCCGCTCCAT 123642
309 eArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGluA 326

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  123641 CCGCCAGTATTACAAGAAGGGCATCATCGGAAGCCAGACATCTCCCAGC 123592

     326 rqLeuValTyrGlnPheValHisProIle 335
         |||||||||||||||||||||||||||||
  123591 GCCTCGTCTACCAGTTCGTGCACCCCATC 123563

seq_name: gb_in2:AF106964

Seq_documentation_block:
LOCUS       AF106964     1832 bp    mRNA            INV       12-MAR-1999
DEFINITION  Stronqylocentrotus purpuratus Ets4 transcription factor (Ets4)
            mRNA, complete cds.
ACCESSION   AF106964
VERSION     AF106964.1  GI:4406283
KEYWORDS    .
SOURCE      Stronqylocentrotus purpuratus.
  ORGANISM  Stronqylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Stronqylocentrotidae; Stronqylocentrotus.
REFERENCE   1  (bases 1 to 1832)
  AUTHORS   Wei,Z., Anqerer,R.C. and Anqerer,L.M.
  TITLE     Identification of a new sea urchin ets protein, SpEts4, by yeast
            one-hybrid screening with the hatching enzyme promoter
  JOURNAL   Mol. Cell. Biol. (1999) In press
REFERENCE   2  (bases 1 to 1832)
  AUTHORS   Wei,Z., Anqerer,R.C. and Anqerer,L.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1998) Biology, University of Rochester, River
            Campus, Rochester, NY 14627, USA
FEATURES             Location/Qualifiers
     source          1. .1832
                     /organism="Stronqylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /dev_staqe="embryo"
     qene            1. .1832
                     /qene="Ets4"
     CDS             134. .1225
                     /qene="Ets4"
                     /codon_start=1
                     /product="Ets4 transcription factor"
                     /protein_id="AAD19942.1"
                     /db_xref="GI:4406284"
                     /translation "MESTGEVCTVSFTPMNDFDATHLAEIMKGSTLLLDLLQQQQKN
                     GPSPSPFSGGFGLPPSPDMQGSFYFPFQNFFPTQQMGSTQQQGGNQTSETEPNEMMRL
                     LMREMTPDTLEDCHKLNLTPNVYCWTPEDVQKWLLCVANEFFICELEMGHFYINGFTI
                     ATLQDVDERHRAPKGGDTLYSVVCTLKSSTHFDVAIQVVTQAQQMMPNTPQPSHFTIG
                     APTTTMMPPKYTYTFVSPVATPGDESGSTGSPSDDETSTPSFAPSGPNTSMTPNHFGG
                     TQTWQFTKFTTTQPNSYSYCTRWTDFTQGTFKTFDSVPVAPTWGTRKNRPAMNYDKLS
                     RSTRQYYRKGTMKKTEVSQRLVYQFVTPC"
BASE COUNT       528 a    448 c    379 q    477 t
ORIGIN


alignment_scores:
            Quality:  487.50            Length:    886
              Ratio:  2.310              Gaps:      9
 Percent Similarity:  54.664   Percent Identity: 30.829

alignment_block:
 US 09-126-945B-2 x AF106964   ..

 Align seq 1/1  to: AF106964  from: 1  to: 1832

       8 LeuSerSerValSerProSerHisLeuLeuLeuProProAspThrValSe 24
         |||:::::::::|||   :::     ||||||     |||   :::::
     200 CTAGCTGAAATCAGCATGGGAGCACGCTTCTAGACCTCGATCTTATCCA 249

      24 rArqThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArqA 41
         ::::                                        :::|||:
     250 ACAACAG.................................CAACGCA 263

      41 rqAspTrpSerProSerPro.....................ProAlaThr 50
         ::      ||||||||||||                    |||   :::
     264 ACGGCCCCTCCCCCAGTCCCTTCAGCGGCGGCTTCGGACTCCCCCCTTCA 313

      51 ProGlu...GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLe 66
         |||:::   ||||||                              ::
     314 CCAGACATGCAAGGG.............................TCCTT 343

      66 nTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerA 83
         :|||||||||:::::::   :::   ::::::     ||:::::::   :
     344 CTATCCAGAGCAGAACTTCTTCCTACACAGCAGATGGGCTCTACCATTC 393

      83 rqGluGluProProGluGluProGluGlnCysProValIleAspSerGlu 99
         ::::::::
     394 AGCAGCAG..................................... 401

     100 AlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGl 116
                                     ||||||     :::::::|
     402 ..........................GGGGGAAACCAGATCTCAGA 421

     116 nHisSerLeuGluGlnValGlnSerMetValValGlyGluValLeuLysA 133
         |   :::   ::::::::::::   :::::::::   ||||::::::   |
     422 GATCGAGGAGAACGACATGATGCGATTACTAATGAGGGAGATGATTCCTG 471

     133 spIleGluThrAlaCysGlyLeuLeuAsnIleThrAlaAspProMetAsp 149
         |||||      |||      ||||||:::      :::
     472 ATATCCTCGAGGATTGCCACAAACTCAACTTGATTCCAAATGTTTACTGC 521

     150 TrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnLy 166
         |||:::|||:::::::|||||||||||||||||||   :::   ::::::::
     522 TGGACGCCGGAAGATGTCCAGAAGTGGTTGTTATGCGTCGCCAATCGCTT 571

     166 rArqLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluL 183
         ::::|||     :::      :::          :::   |||      |
     572 TGAGCTAGGGGAACTCGACATGCGCCACTTTACATCAATGGACCAACCC 621

     183 euCysAlaMetSerGlnGluGlnPheArqGlnArqSerProLeuGlyGly 199
         ||:::::::::::::::   :::|||||||:::|||:::|||      |||
     622 TTGCAACACTCCAGGACGTGGATTCAGGCACAGGGCGCCGAAATGCGGC 671

     200 AspValLeuHisAla.........................HisLeuAs 207
         |||:::|||:::::                         |||:::||
     672 GACATCCTTTACAGCGTTGTATGTCTTCTTAAATCAAGTATTCACTTGA 721

     207 pIleTrpLysSerAlaAlaTrpMetLysGluArqThrSerProGly..... 222
         |:::      :::::::::   :::   ::::::       ||:::
     722 CGTAGCACCCCAGGTCGTACCTCAAGCACAACAGATGATGCCCAATATTC 771

     223 .....AlaIleHisTyr................................ 226
              |||:::
     772 CCCAACCCTCGCACTTCATTACCGGGGCACCCACCACAACCATGATGCCA 821

     226 ..................................................... 226

     822 CCCAAGTACACCTACACGCCCGTCTCACCGGTAGCAACGCCAGGAGACGA 871

     227 .CysAlaSerThrSerGluGluSerTrpThrAspSerGluValAspSerS 243
         ::::::::::::::   |||      :::   ::::::   |||
     872 GAGCGGTAGCACTGGCTCGCCAAGCGATGACGAGATCTCAATCCCATCAC 921

     243 erCysSerGlyGlnPro.............................. 248
         :::   :::   |||
     922 CCGCCCCCAGCGGACCCAACACCTCCATGACCCCGAACCACACAGGAGGT 971

     249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
         |||:::|||||||||||||||||||||||||||||||:::|||:::|||||
     972 ATCCAGCTCTGGCAGTTCTTGAAGGAGCTCCTCCTCAAGCCCAACTCCTA 1021

     265 rGlyArqPheIleArqTrpLeuAsnLysGluLysGlyIlePheLysIleG 282
         |:::      ||||||||||:::::::   :::|||||||||||||||
    1022 CAGCTACTGCATCAGGTGGATAGACAGAACACAAGGTATCTTCAAGATCG 1071

     282 luAspSerAlaGlnValAlaArqLeuTrpGlyIleArqLysAsnArqPro 298
```

```
          |||||||||:::::|||||||||||||||||||:::|||||||||||||||
     1062 AGGACTCTGTCGAGGTCGCCAGACTTTGGGCCTTCAGGAAGAACAGGCCG 1111

      299 AlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLy 345
          ||||||||||||||||||||||||||||||||||||||||||||||:::||
     1112 GCCATGAACTACGACAAGCTCTCTCGCTCCATCAGGCAGTACTACCGCAA 1161

      315 sGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheV 332
          |||||||:::::|||   :::::|||||||||||||||||||||||||
     1162 AGGCATCATGAAGAAGACCGAGGTTTCGCAGCGACTCGTCTACCAGTTTG 1211

      332 alHisPro 334
          ||   |||
     1212 TCACCCCC 1219

seq_name: qb_htg4:AC015474

seq_documentation_block:
LOCUS       AC015474     38178 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015474
VERSION     AC015474.1  GI:6435961
KEYWORDS    HTG; HTGS PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pteryqota; Neoptera; Endopteryqota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 38178)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16 NOV 1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213681 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..38178
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    10464 a   8634 c   8939 q  10141 t
ORIGIN


alignment_scores:
            Quality:  471.00            Length:    287
              Ratio:  2.854               Gaps:      8
 Percent Similarity:  45.296   Percent Identity: 33.798

alignment_block:
 DS:09-126-945B-2 x AC015474  ..

 Align seq 1/1  to: AC015474  from: 1  to: 38178

      163 GluHisGlnTyr.....................ArgLeuProProMe 171
          |||||||||||||                   |||      ||
    15418 GAGCACCAATACTTCTTCTGCTGCTGCTGTCGGTGCAGCCTTTGCACCAT 15467

      171 tGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerG 198
          |     |||                 :::|||||:::
    15468 GGAGACTGCA.................CATAATTTATGCACTTGGATAG 15499

      188 luGluGlnPhe.................ArgGlnArgSerProLeuGly 198
          ::::::|||                 |||   ||||||      |||
    15500 CCCAAGAATTTCCAATTGATATGCCGTCTAGATTGCGATCTGGCCAAGGA 15549

      198 ..................................................  198

    15550 TTAAGGCTTTGTACCCATTCGAAATGTCTCACTTTCGCCTCTCTTTTGCA 15599

      199 ....GlyAspValLeuHisAlaHisLeuAspIleTrpLysSerAla.... 212
              |||:::::|||||||||:::|||:::|||||||||   |||
    15600 GAGCGGCAGCACGCTGCACGCCCAGTTGGAAATTTGGAAGATGGCCTACG 15649

      212 .................................................. 212

    15650 CGGACCAGCCCGCTCATCAGCAGCATTCCCAGCAGTCTGCGAGCACTGAT 15699

      213 AlaTrpMetLysGlnArgThrSerProGlyAla.................. 223
          |||      :::   :::   |||
    15700 CACTGGCCAGCAAGCTATGCGATGCCCCATTTGGACCTGGACTACAACGA 15749

      223 .................................................. 223

    15750 GGACAGCGAAGGTGAGTTCGATCAGGCGAATACTATTTGCATAGACTTCA 15799

      224 ..............................IleHisTyrCysAlaSerThr 230
                                       :::::|||   :::||||||
    15800 AATTGATTTGATCTTGCAAGTTTGACTGACTCGAATATTCCTCTTCTACA 15849

      231 SerGluGluSerTrpThrAspSerGluValAspSerSerCysSer..... 245
          :::::::::          |||   |||:::|||::::::   :::
    15850 GATGACGAC.........GACATGGAAGCAGATGCACAAGTTGCACCCCT 15890

      245 .................................................. 245

    15891 GAACGGCAGTACCACTTCACCGCCCGCTACCAACGCTTCCAACGGCGGTA 15940

      246 ..............................................GlyGln 247
                                                       |||:::
    15941 CGGCGACAGTAAAACGACCGAATGGTGGACGGACCGGCGGTGGGGGCTCC 15990

      248 ProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSe 264
          ||||||||||||||||||||||||||||||||||   :::||||::
    15991 CACATCCACCTGTGGCAGTTCCTCAAGGAGCTGCTGGCTTCGCCCCAGGT 16040

      264 rTyrGlyArgPheIleArgTrpLeuAsnLysGlnLysGlyIlePheLys1 281
          |||       |||||||||:::::::::::|||||||||||||||||
    16041 GAACGGCACAGCCATCCGGTGGATCGACCGAAGCAAGGGCATCTTCAAGA 16090

      281 IeGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
          ||||||||||||:::::|||||:::|||||||||   ||||||||||||
    16091 TCGAGGACTCGGTGAGGGTGGCCAAGCTGTGGGGTCGGCGCAAGAACCGA 16140

      298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 314
          ||||||||||||||||||||||||||||||||||||||||||||||||||
    16141 CCGGCGATGAACTATGATAAGTTGTCCCGCTCCATCAGGCAGTACTACAA 16190

      314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
          ||||||||||:::::|||   :::   |||||||||||||||||||||
    16191 GAAGGGAATCATGAAGAAGACGGAGAGGTCACAGCGGCTGGTCTACCAGT 16240

      331 heValHisPro 334
          ||   ||||||
    16241 TCTGCCATCCC 16251

seq_name: qb_in1:AC008253

seq_documentation_block:
LOCUS       AC008253   180887 bp    DNA             INV       21-MAR-2001
DEFINITION  Drosophila melanogaster, chromosome 3R, region 98B1-98B2, BAC clone
            BACR48A16, complete sequence.
ACCESSION   AC008253
VERSION     AC008253.8  GI:13899341
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pteryqota; Neoptera; Endopteryqota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 180887)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
```

```
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E.,
            Rogers,Y.-H., Hu,M., Baldwin,D., Ramzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Fartan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,J.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nguyen,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 3R, region 98B1-98B2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180887)
  AUTHORS   Celniker,S.E., Ambayani,A., Arcaina,I.I., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Fartan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Humanustl,S.K., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Maida,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Shih,E.,
            Svirskas,R.R., Wan,K.H., Weinburn,I., Zhang,P., Zideman,D.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 21, 2001 this sequence version replaced gi:6633926.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/seq_tools/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1..180887
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="98B1-98B2"
                     /clone="BACR48A16 (D558)"
                     /clone_lib "RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    52999 a  39693 c  39212 g  48983 t
ORIGIN


alignment_scores:
            Quality:  471.00              Length:     287
              Ratio:  2.854                 Gaps:       8
 Percent Similarity:  45.296   Percent Identity: 33.798

alignment_block:
 US-09-126-945B-2 x AC008253  ..

 Align seq 1/1  to: AC008253  from: 1  to: 180887

     163 GluHisGlnTyr.....ArgLeuProProMe 171
         |||||||||||||                 |||       ||
   24263 GACCACCAATACTTCTTGTGCTGCTGCTGCGGTGCAGCGTTTGCACCAT 24312

     171 tGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerG 188
         |    |||          ::::||||||:::
   24313 GGACACTGCA.....CATAATITAIGCACTTGGATAG 24444

     188 InGlnGlnPhe.....ArgGlnArgSerProLeuGly 198
         ::::::|||          |||   ||||||      |||

   24445 GGGAAGAATTGGAATGATATGGTATAATGGTATGTGGTAAGGA 24494

     198 ..................................................  198

   24495 TTAAGGCTTTGTACGCATTCGAAATGTCTCACTTTCGCCTCTCTTTTGCA 24444

     199 ....GlyAspValLeuHisAlaHisLeuAspIleTrpLysSerAla....  212
             |||::::::||||||||:::||||:|||||||||    |||

   24445 GAGCGGCAGCACGCTGCACGCCCAGTTGGAAATTTGGAAGATGGCCTACG 24494

     212                                                  212

   24495 CGGACCAGCCCGCTCATCAGCAGCATTCCCAGCAGTCTGCGAGCACTGAT 24544

     213 AlaArgMetLysGlnArgThrSerProProGlyAla             224
         |||      :::   :::   |||
   24545 CACTGGCCAGCAAGCTATGCGATGCCCCATTTGGACCTGGACTACAACGA 24594

     224 ..................................................  223

   24595 GGACAGCGAACGCTCACTTCGATCAGCCGAATACTATTTGCATAGACTTCA 24644

     224 ..............................IleHisTyrCysAlaSerTh  230
                                      :::::|||   :::||||||
   24645 AATTGATTTGATCTTGGAAGTTTGACTGACCGAATATTCCTCTTCTACA 24694

     231 SerGlnGlnSerArgPheArgAspSerGluValAspSerTerCysSer.....  245
         :::::::::          |||   ||:::||||::::::   :::
   24695 GATGACGAC.....GACATGGAAGCAGATGCACAATTTGCACCCT 24745

     245 ..................................................  245

   24746 GAACGGCAGTACCACTTCACGCCCCGCTACCAACGCTTCCAACGGCCCTA 24795

     246 .....................................................GlyGlu 247
                                                            |||:::
   24796 CCCCGACACTAAAATAACCGAATCTTTCACGCACCCCCCTTCCCCCCCT 24845

     248 ProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSe  264
         ||||||||||||||||||||||||||||||||||||   :::|||:::
   24846 CACATCCACCTGTGGCAGTTCCTCAAGGAGCTGCTGCTTCGCCCACGT 24895

     264 rTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLys  281
         |||     ||||||||||::::::::::||||||||||||||||||
   24896 GAACGGCACAGCCATCCGGTGGATCGACGAAGCAAGGGCATCTTCAAGA 24945

     281 InGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
         ||||||||||||:::::||||||::||||||||   ||||||||||||
   24946 TCGAGGACTCGGTGAGGGTGGCCAAGCTGTGGGGTCGGCGCAAGAACGGA 24995

     298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrTy 314
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   24996 CCGGCGATGAACTATGATAAGTTGTCCCGCTCCATCAGGCAGTACTACAA 25045

     314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
         ||||||||||:::::|||   :::   ||||||||||||||||||||||
   25046 GAAGGAATCATCAAGAAGCACCCAGACATCTCACAGCGTCTGGTCTACCAA 25095

     331 heValHisPro 344
         ||   ||||||
   25096 TCTGCCATCCC 25096

seq_name: gb_in1:AC007818

seq_documentation_block:
LOCUS       AC007818   180549 bp    DNA             INV       06-MAR-2001
DEFINITION  Drosophila melanogaster, chromosome 3R, region 98B-98B, BAC clone
            BACR02M05, complete sequence.
ACCESSION   AC007818
VERSION     AC007818.7  GI:13246626
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
```

```
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 188549)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Booson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garq,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibeqwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Mosbreti,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paraqas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavonq,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Stronq,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencinq of Drosophila chromosome 3R, reqion 98B-98B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188549)
  AUTHORS   Celniker,S.E., Aqbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.E., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshreti,A.R., Moshreti,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinbarq,T., Zhanq,E., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS-64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 6, 2001 this sequence version replaced qi:6957919.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled usinq end sequences from a whole qenome
            shotqun and from subclones of this BAC and its neighborinq clones.
            For further information about this sequence, includinq its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.orq/sequence/) or send email
            to bdqp@fruitfly.berkeley.edu.
FEATURES              Location/Qualifiers
     source           1..188549
                      /orqanism="Drosophila melanoqaster"
                      /strain="y; cn bw sp"
                      /db_xref="taxon:7227"
                      /chromosome="3R"
                      /map="98B-98B"
                      /clone="BACR02M05 (D785)"
                      /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                      Drosophila melanoqaster BAC library, partial EcoRI in
                      pBACe3.6)"
BASE COUNT    51434 a   44300 c   44641 q   50174 t
ORIGIN


alignment scores:
            Quality:   421.00              Length:    287
              Ratio:   2.854                 Gaps:      8
 Percent Similarity:   45.296   Percent Identity:   33.798

alignment block:
 US-09-126-945B-2 x AC007818   ..

 Align seq 1/1  to: AC007818  from: 1  to: 188549

      163 GluHisGlaTyr.....................ArqLeuProProMe 171
          ||||||||||||                    |||      ||
   112802 GAGCACCAATACTTCTTCTGCTGCTGCTGTCGGTGCAGCCTTTGCAGCAT 112851

      171 tGlyLysAlaPheClnGluLeuAlaGlyLysGluLeuCysAlaMetSerG 189
```

```
          |      |||                    :::|||||||:::
   112852 GCAGACTGCA.................CATAATTTATGCACTTGGATAG 112883

      188 luGluGlnPhe.................ArqGlnArqSerProLeuGly 198
            ::::::|||                 |||   ||||||     |||
   112884 CCCAAGAATTTCCAATTGATATGCCGTCTAGATTGCGATCTGGCCAAGGA 112933

      198 ..................................................... 198

   112934 TTAAGGCTTTGTACCCATTCGAAATGTCTCACTTTCGCCTCTCTTTTGCA 112983

      199 ....GlyAspValLeuHisAlaHisLeuAspIleTrpLysSerAla.... 212
              |||:::::::|||||||||||::|||:::|||||||||   |||
   112984 GAGCGGCAGCACGCTGCACGCCCAGTTGGAAATTTGGAAGATGGCCTACG 113033

      212 ..................................................... 212

   113034 CGGACCAGCCCGCTCATCAGCAGCATTCCCAGCAGTCTGCGAGCACTGAT 113083

      213 AlaTrpMetLysGluArqThrSerProGlyAla.................. 223
            |||      :::   :::   |||
   113084 CACTGGCCAGCAAGCTATGCGATGCCCCATTTGGACCTGGACTACAACGA 113133

      223 ..................................................... 223

   113134 GGACAGCGAAGGTGAGTTCGATCAGCCGAATACTATTTGCATAGACTTCA 113183

      224 ..............................IleHisTyrCysAlaSerPhe 240
                                       :::::::|||   :::|||||||
   113184 AATTGATTTGATCTTGGAAGTTTGACTGACTCGAATATTCCTCTTCTACA 113233

      241 SerGluGluSerTrpThrAspSerGluValAspSerSerCysSer..... 245
          :::::::::          |||   |||:::|||::::::   :::
   113234 GATGACGAC.........GACATGGAAGCAGATGCACAAGTTGCACCCCT 113274

      245 ..................................................... 245

   113275 GAACGGCAGTACCACTTCACCGCCCGCTACCAACGCTTCCAACGGCGGTA 113324

      246 ...............................................GlyGln 247
                                                        ||::::
   113325 CGGCGACAGTAAAACGACCGAATGGTGGACGGACCGGCGGTGGGGGCTCC 113374

      248 ProIleHisLeuTrpGluPheLeuLysGluLeuLeuLeuLysProHisSe 264
          ||||||||||||||||||||||||||||||||||||   :::|||:::
   113375 CACATCCACCTGTGGCAGTTCCTCAAGGAGCTGCTGGCTTCGCCCCAGGT 113424

      264 rTyrGlyArqPheIleArqTrpLeuAsnLysGluLysGlyIlePheLys! 281
          |||      ||||||||||:::::::::|||||||||||||||||||
   113425 GAACGGCACAGCCATCCGGTGGATCGACCGAAGCAAGGGCATCTTCAAGA 113474

      281 leGluAspSerAlaGlnValAlaArqLeuTrpGlyIleArqLysAsnArq 297
          ||||||||||:::::|||||::|||||||||   ||||||||||||
   113475 TCGAGGACTCGGTGAGGGTGGCCAAGCTGTGGGGTCGGCGCAAGAACCGA 113524

      298 ProAlaMetAsnTyrAspLysLeuSerArqSerIleArqGlaTyrTyrLy 314
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   113525 CCGGCGATGAACTATGATAAGTTGTCCCGCTCCATCAGGCAGTACTACAA 113574

      314 sLysGlyIleIleArqLysProAspIleSerGluArqLeuValTyrGluP 331
          ||||||||||:::::|||   :::   ||||||||||||||||||||||
   113575 GAAGGGAATCATCGAAGAAGACGGAGAGGTCACAGCGGCTGGTCTACCAGT 113624

      331 heValHisPro 334
          ||   ||||||
   113625 TCTGCCATCCC 113635

seq_name: qb_inl:AC008217

seq_documentation_block:
LOCUS       AC038217   193245 bp    DNA             INV       21-MAR-2001
DEFINITION  Drosophila melanoqaster, chromosome 3R, reqion 98B-98B, BAC clone
            BACR10J03, complete sequence.
```

```
ACCESSION   AC008217
VERSION     AC008217.5  GI:13399346
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 193245)
  AUTHORS   Celniker,S.E., Adams,M.D., Krommiller,R., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Cocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Drosnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,I.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,E., Moshreti,A.,
            McIntosh,I.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paraqas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phonanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,E., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 3R, region 98B-98B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193245)
  AUTHORS   Celniker,S.E., Aubayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.F., Karta,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Maeda,P.,
            Moshreti,A.F., Moshreti,M., Nixou,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Suit,E.,
            Svirskas,R.F., Wan,K.H., Weinburg,E., Zhang,P., Zitzer,J.F. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 21, 2001 this sequence version replaced gi:6957918.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1..193245
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="98B-98B"
                     /clone="BACR10J03 (D786)"
                     /clone_lib="BDGP 98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    55746 a  43163 c  42757 g  51589 t
ORIGIN


alignment_scores:
            Quality:   471.00              Length:     287
              Ratio:   2.854                Gaps:       8
 Percent Similarity:   45.296   Percent Identity:  33.798

alignment_block:
 DS-09-126-945B-2.x AC008217  ..

 Align seq 1/1  to: AC008217  from: 1  to: 193245
```

```
  163 GluHisGlnTyr.....ArgLeuProProMe 171
      ||||||||||||                ||| ||
55525 GAGCACCAATACTTCTTCTGCTGCTGTCGGTGCAGCCTTTGCACCAT 55574

  171 GlyLysAlaPheGluGluLeuAlaGlyLysGluLeuCysAlaMetSerG 188
      |   |||                 ::::||||||||:
55575 GGAGACTGCA.................CATAATTTATGCACTTGGATAG 55606

  188 LuGluGlnPhe.....ArgGlnArgSerProLeuGly 198
      :::::|||                |||  ||||||  |||
55607 CCCAAGAATTTGCAATTGATATGCCGTCTAGATTGCCATCTGGCCAAGGA 55656

  198 ..... 198

55657 TTAAGGCTTTGTACCCATTCGAAATGTCTCACTTTGCCTCTCTTTTGCA 55706

  199 ....GlyAspValLeuHisAlaHisLeuAspIleTrpLysSerAla.... 212
          |||:::::|||||||||:::|||:::|||||||||   |||
55707 GAGCGGCAGCACGTGCACGCCCAGTTGGAAATTTGGAAGATGGCCTACG 55756

  212 ..... 212

55757 CGGACCAGCCCGCTCATCAGCAGCATTCCCAGCAGTCTGCGAGCACTGAT 55806

  213 AlaTrpMetLysGluArgThrSerProGlyAla..... 223
      |||       :::   :::   |||
55807 CACTGGCCAGCAAGCTATGCGATGCCCCATTTGGACCTGGACTACAACGA 55856

  223 ..... 223

55857 GGACAGCGAAGTGAGTTCGATCAGGCGAATACTATTTGCATAGACTTCA 55906

  224                                 IleGlyTyrCysAlaSerThr 230
                                      :::::|||   :::||||||
55907 AATTGATTTGATCTTGGAAGTTTGACTGACTCGAATATTCCTCTTCTACA 55956

  231 SerGluGluSerTrpThrAspSerGluValAspSerSerCysSer       245
      :::::::::          |||   |||:::|||:::::::   :::
55957 GATGACGAC.........GACATGGAAGCAGATGCACAAGTTGCACCCCT 55997

  245 ..... 245

55998 CAACGGCAGTACCACTTCACCGCCGCTACCAACGCTTCCAACGGCGGTA 56047

  246 .....GlyGln 247
                ||:::
56048 CGGCGACAGTAAAACGACCGAATGGTGGACCGACCGGCGGTCGGGGCTCC 56097

  248 ProIleHisLeuTrpGluPheLysGlyGlnLeuLeuLysProHisSe 264
      ||||||||||||||||||||||||||||||||   ::::||:::
56098 CACATCCACCTGTGGCAGTTCCTCAAGGAGCTGCTGGCTCCGCCCAGGT 56147

  264 TyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysI 281
        |||       ||||||||||:::::::::::|||||||||||||||
56148 GAACGGCACAGCCATCCGGTGGATCGACGGAAGCAAGGGCATCTTCAACA 56197

  281 IeGluAspSerAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArg 297
      |||||||||||:::::|||||||:::|||||||||   ||||||||||
56198 TCGAGGACTCGGTGAGGGTGGCCAAGCTGTGGGGTCGGCGCAAGAACGGA 56247

  298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyrTyrLy 314
      ||||||||||||||||||||||||||||||||||||||||||||||||||
56248 CCGGCGATGAACTATGATAAGTTGTCCCGCTCCATCAGGCAGTACTACAA 56297

  314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
      ||||||||||:::::|||   :::   |||||||||||||||||||||
56298 GAAGGGAATCATGAAGAAGACGGAGAGGTCACAGCGGCTGGTCTACCAGT 56347

  331 heValHisPro 334
      ||   ||||||
56348 TCTGCCATCCC 56358

seq_name: qb_in2:AE003762
```

```
seq_documentation_block:
LOCUS       AE003762    230940 bp    DNA              INV       07-JAN-2003
DEFINITION  Drosophila melanogaster genomic scaffold 142000013486035 section 87
            of 105, complete sequence.
ACCESSION   AE003762 AE002708
VERSION     AE003762.2  GI:10726810
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 230940)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D.C., Scheeler,F., Shen,H., Shue,B.C., Sidén-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 230940)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7301615.
FEATURES             Location/Qualifiers
     source          1..230940
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
     mRNA            join(<2097..2414,3389..3409,3466..3802,3854..4153,
                     4211..4348,4402..4715,4773..4972,5037..5218,5408..5512,
                     6082..6217,6338..6773,6844..7041,7442..7722,7794..7889,
                     7946..8563,8624..9356,9414..9551,9612..9910,9967..10145,
                     10199..10805,10863..11420,11477..11780,11842..12791,
                     12847..14304,14685..14853,14914..15248,15446..15574,
                     15646..15869)
                     /gene="CG18447"
                     /product="CT41992"
                     /db_xref="FLYBASE:FBan0018447"
                     /db_xref="FLYBASE:FBgn0039536"
                     /evidence=not_experimental
     gene            <2097..>15869
                     /gene="CG18447"
                     /map="98A6-98A8"
                     /db_xref="FLYBASE:FBan0018447"
                     /db_xref="FLYBASE:FBgn0039536"
                     /evidence=not_experimental
     CDS             join(2097..2414,3389..3409,3466..3802,3854..4153)
                     /gene="CG18447"
                     /note="CG18447 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0018447"
                     /db_xref="FLYBASE:FBgn0039536"
                     /evidence=not_experimental
                     /protein_id="AAF56734.1"
                     /db_xref="GI:7301616"
                     /translation="MNCSAQPAVARHLICVLVPFQLATMVTNAAGIAAIGGATSNLIN
                     NNNIQTNNNSHGANNNNDDFDFDQQSGLQHLGLPVSVQTFLWRQIAPFIKPKLCKLHE
                     SICLESNLACKSCEKVLVQNIQRGLSPELRNGIAVLCWKLLQGALPHVMHACAALLY
                     NRVFFMQAIGPVFIELLYIMQWILLYAAEFCADDFGGIFDLGIGDAAEIFKSKSIDQYLF
                     SVPLTITVSLCVPICTYYTPPKGIGFPKLSPGEWH"
     mRNA            16549..18002
                     /gene="CG5590"
                     /product="CT17678"
                     /db_xref="FLYBASE:FBan0005590"
                     /db_xref="FLYBASE:FBgn0039537"
     gene            16549..18992
                     /gene="CG5590"
                     /map="98A6-98A8"
                     /db_xref="FLYBASE:FBan0005590"
                     /db_xref="FLYBASE:FBgn0039537"
     CDS             16658..17896
                     /gene="CG5590"
                     /note="CG5590 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0005590"
                     /db_xref="FLYBASE:FBgn0039537"
                     /protein_id="AAF56735.1"
                     /db_xref="GI:7301617"
                     /translation="MIN[illegible]
                     PHPKLAGLLYSAAEKLEKAGGKAYPVVPVSPEQVESAVEAAVAKEGGLDIVINNAS
                     AISLLNTPSQMKEKYELMHINTRGTFLVSKEVVPLMEKSNHAHILNISPPLSMKEEW
                     FGPHVAYTMAKYGMSMCVLGMAAEFKDEGISVNALWPRTAIHTAATEMLIGPDSAKWS
                     RKPEIMADAAYAILTREPPQSTGQFFVDDEVLESAGITDLTEYACTPENADKLMVDFF
                     VEEKGAPVENSAAAEDAAAPASGGDVKIPQLERKIESLESPEIVSKIQAVEQENISGA
                     LQSIWIDLENGSIGGGAGTLFIAATLAILLMGSKNLIDMFSGELEAAIAYMICEEIS
                     GDPQKALALEKLMKALKSKL"
     mRNA            join(<18484..18404,18475..>18927)
                     /gene="CG12883"
                     /product="CT42027"
                     /db_xref="FLYBASE:FBan0012883"
                     /db_xref="FLYBASE:FBgn0039538"
                     /evidence=not_experimental
     gene            <18484..>18927
                     /gene="CG12883"
                     /map="98A6-98A8"
                     /db_xref="FLYBASE:FBan0012883"
                     /db_xref="FLYBASE:FBgn0039538"
                     /evidence=not_experimental
     CDS             join(18484..18404,18475..18927)
                     /gene="CG12883"
                     /note="CG12883 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012883"
                     /db_xref="FLYBASE:FBgn0039538"
                     /evidence=not_experimental
                     /protein_id="AAF56736.1"
                     /db_xref="GI:7301618"
                     /translation="MASSSSKNKTLQLIRCFCGCSNMIVAEVRKIYVLSNSVHEILLD
```

```
                     AEATRLLRKFMFTRKSGDKDFAFQYLDTYEGCAEFLAFHQRIFTQDEVDELVDLGLDY
                     DLEQDLSRRMLSADRTDISSGLYRTQSKCRNETEGSSDFRDFRDATADKMREVPK"
     mRNA            complement(join(19405. .19569,19627. .20979))
                     /gene="Ring"
                     /note "Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product "CT17690"
                     /db_xref="FLYBASE:FBan0005595"
                     /db_xref "FLYBASE:FBgn0020911"
     gene            complement(<19405. .>20979)
                     /gene="Ring"
                     /note="CG5595"
                     /map "98A6-98A8"
                     /db_xref "FLYBASE:FBan0005595"
                     /db_xref "FLYBASE:FBgn0020911"
     CDS             complement(join(19525. .19569,19627. .20889))
                     /gene="Ring"
                     /note "Ring gene product; Nucleotide sequence of the
                     Celera sequence differs from the published sequence for
                     this transcript"
                     /codon_start=1
                     /db_xref "FLYBASE:FBan0005595"
                     /db_xref "FLYBASE:FBgn0020911"
                     /protein_id "AAF56737.1"
                     /db_xref "GI:7301619"
                     /translation "MESIDPAPNKTWFLSLYFLQRKDQEVLLDSTFIAVSPRSLDSEL
                     MCPICLDMLKKTMLTKFCLHRFCSQCIVIAIRSGNEECPVCKELVSFRSLRANPNLR
                     LLISKIYPSREFVFAIQFKVMAKTNQTQSQQALVNSTNEGIKILQSQNRPQPPTTKGG
                     GGGGGGNGNGAANVAAPPAPGAPTAVGRNASNQMHVHDTACNDSNSNTNSTGKLNKL
                     PGHSGTSAASAITSASNAAPSSSANSGASTSATRMQVDDASNIPSVKSTPSFVPSNSS
                     SSFPKPAMSVLISFRSPPSPSPSQMPCPTPCDSNIDTPCDCNCELSINEIILVTFPH
                     PTEMSADNQLIRALKENCVPYIKTTANAIVDRLSKYLAMELTLDLQADLPEACRVLNF
                     CIYVAPQPQQLVTLNGNQTLRQVNDKFWEVNEFMEMEYSwsss"
     mRNA            join(24956. .25239,28760. .28978,29527. .29971,30701. .31087)
                     /gene "CG12880"
                     /product "CT32024"
                     /db_xref "FLYBASE:FBan0012880"
                     /db_xref="FLYBASE:FBgn0039539"
     gene            <24956. .>31087
                     /gene="CG12880"
                     /map="98A6-98A8"
                     /db_xref "FLYBASE:FBan0012880"
                     /db_xref "FLYBASE:FBgn0039539"
     CDS             join(25160. .25239,28760. .28978,29527. .29971,30701. .31087)
                     /gene="CG12880"
                     /note="CG12880 gene product"
                     /codon_start=1
                     /db_xref "FLYBASE:FBan0012880"
                     /db_xref "FLYBASE:FBgn0039539"
                     /protein_id="AAF56738.1"
                     /db_xref="GI:7301620"
                     /translation "MEKVSVGKHKTLNMKFSWMGIAAWLLVLLTAAMVFGASNS
                     VALAAPTASAAKDVAHMHRSKRLDRRSGAGSSPLIAIKQPSQLAKSAHCLNKKLLSQM
                     GFMKVKAPNSHNKKRLAVESRRHSSRDDSHMFTIKLPPNMHYYAGPGAKNSIAEQQDK
                     SGHFGNAPSIMASSTPVAAAASAAVAISPVAEASALKANGALVNGLPLGYLLPSHNAL
                     GLDLADCPAGPWPWPFLALSPVDLGPCGQVSFPFSSNGRPCRIYHWNLPVLKEALAK
```

```
alignment_scores:
            Quality:  371.00              Length:     287
              Ratio:  2.854                 Gaps:       8
 Percent Similarity:  45.296   Percent Identity:  33.798

alignment_block:
 US-09-126-945B-2 x AE003762   ..

 Align seq 1/1  to: AE003762  from: 1  to: 230940

     163 GluHisGlnTyr.....ArgLeuProProMe 171
         ||||||||||||                   ||| |
   85213 GAGCACCAATACTTCTTCTGCTGCTGCTGTCGGTGCAGCCTTTGCACCAT 85262

     171 tGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerG 188
         |   |||                  ::::||||||:::
```

```
   85263 GGAGACTGCA.....................CATAATTTATGCACTTGGATAG 85294

     188 luGluGlnPhe.....................ArgGlnArgSerProLeuGly 198
           :::::::||                    ||   ||||||       |||
   85295 CGCAAGAATTCCAATTGATATGCCGTCTAGATTGCGATCTGGCCAAGGA 85344

     198 ..................................................... 198

   85345 TTAAGGCTTTGTACCCATTCGAAATGTCTCACTTTCGCCTCTCTTTTGCA 85394

     199 ....GlyAspValLeuHisAlaHisGluAspIleTrpLysSerAla.... 212
             |||::::::|||||||||:::||:::|||||||||    |||
   85395 GAGCGGCAGCACGCTGCACGCCCAGTTGGAAATTTGGAAGATGGCCTACG 85444

     212 ..................................................... 212

   85445 CGGACCAGCCCGCTCATCAGCAGCATTCCCAGCAGTCTGCGAGCACTGAT 85494

     213 AlaTrpMetLysGluArgThrSerProGlyAla.................. 223
           |||       :::   :::   |||
   85495 CACTGGCCAGCAAGCTATGCGATGCCCATTTGGACCTGGACTACAAGGA 85544

     223 ..................................................... 223

   85545 GGACAGCGAAGGTGAGTTCGATCAGGCGAATACTATTGCATAGACTTCA 85594

     224 ..............................IleGlnTyrCysAlaSerThr 230
                                      ::::::|||   :::|||||
   85595 AATTGATTTGATCTTGGAAGTTTGACTGACTCGAATATTCCTCTTCTACA 85644

     231 SerGlyGlnSerTrpThrAspSerGluValAspSerSerCysSer..... 245
         ::::::::          |||   ||||:|||:::::   :::
   85645 GATCACGAC.........CACATCGAAGCAGATCACAACTTCGCACCCT 85685

     245 ..................................................... 245

   85686 GAACGGCAGTACCACTTCACCGCCCGCTACCAACGCTTCCAACGGCGGTA 85735

     246 .....................................................GlyGln 247
                                                            ||::::
   85736 CGGCGACAGTAAAACGACCGAATGCTGGACCGACCGGCGCTGGCGCTCC 85785

     248 ProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSe 264
         ||||||||||||||||||||||||||||||||||||   ::::||||::
   85786 CACATCCACCTGTGCCAGTTCCTCAAGGAGCTGCTGCTCAAGCCCCACAGT 85835

     264 rTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysI 281
           |||       |||||||||:::::::::::::|||||||||||||||
   85836 GAACGGCACAGCCATCCGGTGGATCGACCGAAGCAAGGGCATCTTCAAGA 85885

     281 leGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
         ||||||||||:::::|||||::::|||||||   ||||||||||||
   85886 TCGAGGACTCGGTGAGGGTGGCCAAGCTGTGGGGTATCGCCGCAAGAACCGA 85935

     298 ProAlaMetAspTyrAspLysCysArgSerIleArgGlnTyrTyrLy 314
         ||||||||||||||||||||||||||||||||||||||||||||||||
   85936 CCGCCGATGAACTATGATAAGTTGTGCCGCTCCATCAGGCAGTACTACAA 85985

     314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
         ||||||||||:::::|||   :::   ||||||||||||||||||||
   85986 GAAGGGAATCATCAAGAAGCCGGACATCTCACAGCGGCTGGTCTACCAGT 86035

     331 heValHisPro 334
         ||   ||||||
   86036 TCTGCCATCCC 86046

seq_name: gb_in3:DRODETS4A

seq_documentation_block:
LOCUS       DRODETS4A      862 bp    DNA              INV       28 SEP 1999
DEFINITION  Drosophila melanogaster chromosome 3 D-ets-4 DNA binding domain
            protein gene, partial cds.
ACCESSION   M88474
```

```
VERSION     M88474.1  GI:157195
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 362)
  AUTHORS   Chen,T., Bunting,M., Karim,F.D. and Thummel,C.S.
  TITLE     Isolation and characterization of five Drosophila genes that encode
            an ets related DNA binding domain
  JOURNAL   Dev. Biol. 151, 176-191 (1992)
  MEDLINE   92249640
FEATURES             Location/Qualifiers
     source          1..362
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3"
                     /map="3R; 98B"
     mRNA            <1..>345
                     /product="D ets-4 DNA binding domain protein"
     CDS             <1..345
                     /codon_start=1
                     /product="D ets-4 DNA binding domain protein"
                     /protein_id="AAA28451.1"
                     /db_xref="GI:157196"
                     /translation="[illegible]
                     AIRWIDEPEHIEFIIKLVPVAFLWGFRNRFAMRTLR[illegible]
                     RLVYQFCHPYSQ"
BASE COUNT       90 a     99 c    113 g     60 t
ORIGIN


alignment scores:
            Quality:   360.50              Length:     110
              Ratio:     4.144               Gaps:       2
 Percent Similarity:   79.091   Percent Identity: 66.364

alignment block:
 US-09-126-945B-2 x DRODETS4A    ..

 Align seq 1/1  to: DRODETS4A  from: 1  to: 362

      226 TyrCysAlaSerThrSerGlnGluSerTrpThrAspSerGlnValAspSe 242
          |||    |||   :::   |||   ||||||||||   :::
       21 TACGGCGACAGTAAAACGACCGAATGGTGGACGGACCGGCGG........ 62

      242 rSerCysSerGlyGlnPro.IleHisLeuTrpGlnPheLeuLysGluLeu 258
                 |||   ||| ||||||||||||||||||||||||||||||||
       63 .......TGGGGGCTCCCACATCCACCTGTGGCAGTTCCTCAAGGAGCTG 105

      259 LeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGl 275
          |||   :::|||:::       |||       ||||||||||:::::::::
      106 CTGGCTTCGCCCCAGGTGAACGGCACAGCCATCCGGTGGATCGACGAAG 155

      275 uLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpG 292
          :|||||||||||||||||||||||||||:::::|||||::|||||||
      156 CAAGGGCATCTTCAAGATCGAGGACTCGGTGAGGGTGGCCAAGCTGTGGG 205

      292 lyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSer 308
          ||   ||||||||||||||||||||||||||||||||||||||||||||
      206 GTCGGCGCAAGAACCGACCGGCGATGAACTATGATAAGTTGTCCCGCTCC 255

      309 IleArgGluTyrTyrLysLysGlyIleIleArgLysProAspIleSerGl 325
          |||||||||||||||||||||||||||||:::::|||   :::   |||||
      256 ATCAGGCAGTACTACAAGAAGGGAATCATCAAGAAGACGGACGAGGTCACA 305

      325 nArgLeuValTyrGlnPheValHisPro 334
          |||||||||||||||||   ||||||
      306 GCGGCTGGTCTACCAGTTCTGCCATCCC 333

seq_name: qb_ov:XLAJ4126
```

```
seq_documentation_block:
LOCUS       XLAJ4126     2012 bp    mRNA            VRT       21-JAN-2000
DEFINITION  Xenopus laevis erg gene (erg_E).
ACCESSION   AJ224126
VERSION     AJ224126.1  GI:5420047
KEYWORDS    ERG gene; transcription factor.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 2012)
  AUTHORS   Baltzinger,M., Mager-Heckel,A.M. and Remy,P.
  TITLE     Xl erg: expression pattern and overexpression during development
            plead for a role in endothelial cell differentiation
  JOURNAL   Dev. Dyn. 216 (4-5), 420-433 (1999)
  MEDLINE   20099678
REFERENCE   2  (bases 1 to 2012)
  AUTHORS   Baltzinger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1998) Baltzinger M., UPR9005, MMDCD, Centre
            National de la Recherche Scientifique, 15, RUE Rene Descartes,
            67084, STRASBOURG, FRANCE
FEATURES             Location/Qualifiers
     source          1..2012
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="lambda ZAP II erg_E"
                     /clone_lib="lambda ZAP II"
                     /dev_stage="stage 30 embryo"
     gene            225..1595
                     /gene="erg"
     CDS             225..1595
                     /gene="erg"
                     /function="transcription factor"
                     /codon_start=1
                     /protein_id="CAB46567.1"
                     /db_xref="GI:5420048"
                     /db_xref="SPTREMBL:Q9W6Z9"
                     /translation="[illegible]
                     SSPYGQTSKMSPKVPQQHWLSQPPSPVTTKMPCNPSQVNGSPSSPTDPCSTGGGSKMGG
                     GSDNVGMNYGSYIEEKHTSPPNMTTNERRVIVPADPTLWSTDHVRQWLEWAIKEYGLP
                     DVDVLLFQNIDGKELCKMTKEDFHRLTPSYNADILLSHLHYLRERGASFTPPNSSVYQ
                     DANQRIPSRQDLSYEPSRPSAWTNHPAPPSKASQPSTTVPKTEDPRPQTDPYQTLGPT
                     SSWIANPGSQQTQIWQPTIPTTSDSSNSNCTTWPGINGPFKMTDPDEVARRWGERKSK
                     PNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPPESIMYKYPSEL
                     PYMSSYHAHPQKMNFVAPHPPALPVTSSSFFAAPNAYWNSPTGSIYPNTRLPASHMSS
                     HLGTYY"
BASE COUNT      634 a    479 c    895 g    504 t
ORIGIN


alignment scores:
            Quality:   307.50              Length:     432
              Ratio:     1.464               Gaps:        19
 Percent Similarity:   48.611   Percent Identity: 26.157

alignment block:
 US-09-126-945B-2 x XLAJ4126    ..

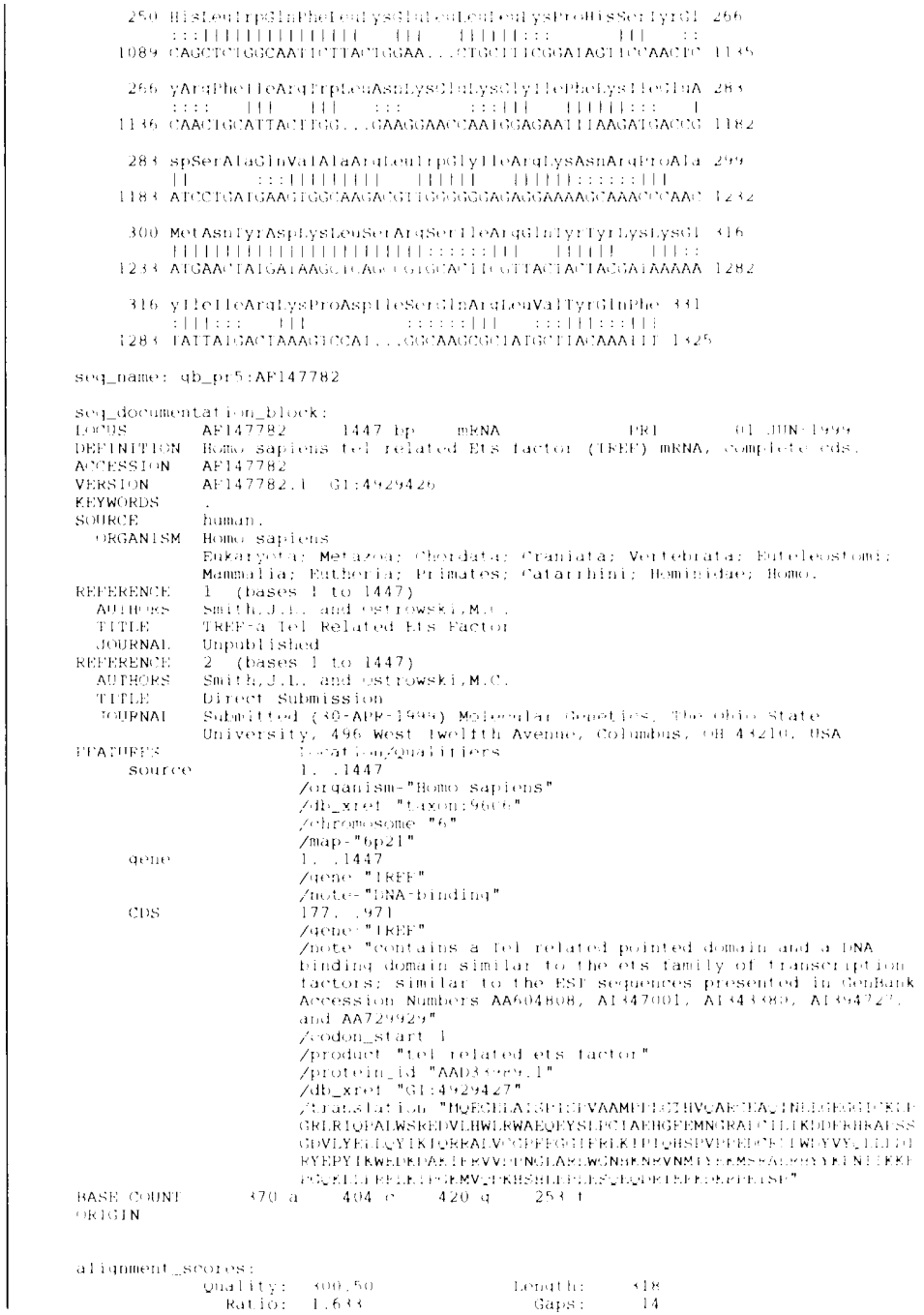
 Align seq 1/1  to: XLAJ4126  from: 1  to: 2012

       8 LeuSerSerValSerProSerHisLeu.....LeuLeuProPr 20
         ||||||||||:   |||     |||     |||   |||
      93 CTATCCACCCTCCTTCCTATTCTGCTGCAGTGGAAAAGTCTCCATCCACG 142

      20 nAspThrValSerArg.....ThrGlyLeuGluL 30
         |||       ||||||                 |||   |||:
     143 GGATCTGTCCTCAGGTCTATTCAGCAGACTCACTATCCTTGGGACTGAAC 192

      30 ysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSer..... 44
         :::::::::::::                 |||||||:::
     193 GTGTGGTATGTAATCCA.....GACTGGAAGAAGGAC 224
```





```
seq_name: qb_pr5:AF147782

seq_documentation_block:
LOCUS       AF147782     1447 bp    mRNA            PRI       01-JUN-1999
DEFINITION  Homo sapiens Tel related Ets factor (TREF) mRNA, complete cds.
ACCESSION   AF147782
VERSION     AF147782.1  GI:4929426
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1447)
  AUTHORS   Smith,J.L. and Ostrowski,M.C.
  TITLE     TREF-a Tel Related Ets Factor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1447)
  AUTHORS   Smith,J.L. and Ostrowski,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-1999) Molecular Genetics, The Ohio State
            University, 496 West Twelfth Avenue, Columbus, OH 43210, USA
FEATURES             Location/Qualifiers
     source          1..1447
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21"
     gene            1..1447
                     /gene="TREF"
                     /note="DNA-binding"
     CDS             177..971
                     /gene="TREF"
                     /note="contains a Tel related pointed domain and a DNA
                     binding domain similar to the ets family of transcription
                     factors; similar to the EST sequences presented in GenBank
                     Accession Numbers AA604808, AI447001, AI443080, AI964727,
                     and AA729929"
                     /codon_start=1
                     /product="tel related ets factor"
                     /protein_id="AAD33999.1"
                     /db_xref="GI:4929427"
                     /translation="MQECELAISPTCPVAAMPPHSZHVCAECEACINELGEGQICECP
                     GRLRIQPALWSKEDVLHWLRWAEQEYSLPCTAEHGFEMNGRALCILIKDDFEGHKAFSS
                     GDVLYECLQYIKTQRKALVCPFEGGIFRELKIPLQHSPVPFPEDCKILWLYVY_IIIDI
                     RYEPYIKWEDKDAEIPPVVPPNGLAAELWGNSKNPVNMIYSAMSSALRQYYLENIIKKI
                     PQLELIPEELKIPGMVLPKHSHLIPGLSLLQQDKIEEIQRQPEISF"
BASE COUNT       370 a     404 c     420 g     253 t
ORIGIN


alignment_scores:
            Quality:  300.50              Length:     518
              Ratio:  1.633                 Gaps:      14
```

```
 Percent Similarity:  57.862   Percent Identity: 28.616

Alignment block:
 US-09-126-945B-2 x AF147782   ..

 Align seq 1/1  to: AF147782  from: 1  to: 1447

       86 ValGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGl  52
          :::|||   :::|||:::   :::   |||:::   |||:::::::   ::
           9 CTTGGCGCTTCTCGAAAGAGATTCCTCCCACGGCGACCTTCCAGTTCTCG  58

       52 a.........GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetL  66
          :         :::|||||||:::
       59 GAGCCAGGTTAGCGGTTTGGCG.............................  80

       66 euTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSer  82
                     :::::::::   :::|||:::|||   |||   ||||||
       81 ...........GAGGAGGACTGCGGGGCGCGGGCCTAGGGCCCCAGCAGC  119

       83 ArgGluGluProProGluGluProGluGlnCysProValIleAspSerGl  99
          :::   :::          :::   ::::::   ||||||   ::::||
      120 CACGGCCAGGGGAGCGCTCAA...GACAGAAAGCCGGTGGCTTCCTCACC  166

       99 nAlaProAla.........GlySerLeu.....................  105
               |||:::         |||:::|||
      167 TCCACCTGTAATGCAGGAGGGAGAATTGGCTATTTCTCCTATAAGCCCTG  216

      106 .....AspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGlu  120
           :::   |||   |||       ::::::   :::
      217 TGGCAGCCATGCCTCCCCTAGGCACCCACGTGCAAGCCAGATGTGAAGCT  266

      121 GlnValGlnSerMetValValGlyGluValLeuLysAspIleGluThrAl  137
          ||||::::::   :::       |||   :::
      267 CAAATTAACCTGCTGGGTGAAGGGGGATC.....................  296

      137 aCysLysLeuLeu...AsnIleThrAlaAspProMetAspTrpSerProS  153
           |||||||||      ::::::      :::|||      ||||||   :
      297 .TGCAAGCTGCCAGGAAGACTCCGCATCCAGCCGCACTGTGGAGCAGGG  345

      153 crAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuPro  169
          :::::|||      ||||||   |||:::|||:::::|||   ||||||
      346 AGCACGTGCTGCACTGGCTGCGCTGGGCAGAGCAGGAGTACTCTCTGCCA  395

      170 ProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMe  186
          :::::      |||:::   |||:::   ||||||   ::
      396 TGCACCGCGGAGCACGGTTCGAGATGAACGGACGCGCCCTCTGCATCCT  445

      186 tSerGluGluGluPheArgGluArgSerProLeuGlyGlyAspValLeuH  202
          ::::::::::::|||||:::|||:::|||   :::|||||||||||:
      446 CACCAAGGACGACTTCCGGCACCGTGCGCCCAGCTCAGGTGACGTCCTGT  495

      203 isAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThr  219
          ::      |||:::      |||:::              ::::::|||:::
      496 ATGAGCTGCTCCAGTACATCAAGACC............CAGCGGCGAGCC  543

      220 .........SerPro......GlyAlaIleHisTyrCysAlaSerThrSe  241
                  :::|||      |||:::|||             :::   ::
      544 CTGGTGTGTGGACCCTTTTTGGAGGGATCTTCAGGCTGAAGACGCCCAC  593

      231 rGluGluSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnP  248
          ::::::::|||   :::       |||           ::::||
      594 CCAGCACTCTCCCAGTCCCCCCGGAA.........GACTGCCGC.......  617

      248 roIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSer  264
                |||||:::::::::   :::||||||||||       :::
      618 .....CTGCTGTGGGATTACGTGTATCAGCTGCTCCTTGATACCCGA...  659

      265 TyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIl  281
          |||     :::|||:::|||   :::|||:::      ||||||::::
      660 TATGAGCCCTACATCAACTGGGAAGACAAGGACGCCAAGATCTTCCGAGT  709

      281 eGluAspSerAlaGlnValIleArgLeuTrpGlyIleArgLysAsnArgP  298
          :   |||         :::||||||||||||||   :::||||||||||
```

```
      710 tGTGGATCCAAATGGGCTCGCCAGACTCTGGGGAAATCACAAGAACCGGG  759

      298 roAlaMetAsnLysAspLysLeuSerArgSerIleArgGlnLysTyrLys  314
               |||:::|||:::|||:::||||||:::::|||:::||||||||||
      760 TGAACATGACCTACGAGAACATGTCTCGTGCCCTGCGCCACTATTATAAG  809

      315 LysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPh  331
                ::::|||||||:|||   :::   :::|||:::|||:::::::||
      810 CTTAATATCATTAAGAAG...GAACCGGGGCAGAAACTCCTGTTTCAGATT  856

      331 eVal 332
          |:::
      857 TCTA 860

Seq_name: qb|prl0:HUMERG2

Seq_documentation block:
LOCUS        HUMERG2       3166 bp ss-RNA             PRI       08-NOV-1994
DEFINITION   Human erg2 gene encoding erg2 protein, complete cds.
ACCESSION    M17254
VERSION      M17254.1  GI:182186
KEYWORDS     erg 2 protein.
SOURCE       Human cell line COLO 320, cDNA to mRNA, clone lambda 12.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 3166)
  AUTHORS    Rao,V.N., Papas,T.S. and Reddy,E.S.
  TITLE      erg, a human ets related gene on chromosome 21: alternative
             splicing, polyadenylation, and translation
  JOURNAL    Science 237 (4815), 635-639 (1987)
  MEDLINE    87263429
FEATURES             Location/Qualifiers
     source          1..3166
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="21q22.3"
     gene            257..1645
                     /gene="ERG"
     CDS             257..1645
                     /gene="ERG"
                     /note="erg 2 protein"
                     /codon_start=1
                     /db_xref="GDB:G00-119-884"
                     /protein_id="AAA52398.1"
                     /db_xref="GI:182187"
                     /translation="MEQIVPFPAAHEEEALSVVSEEQSLEFCAYGTPRLANTEMTASS
                     SSDYGQDFKMGFPEVPQQHWLRQPPARVTTKMDFNPPQVNFCSRNSPDFPCTVAECCKMVG
                     SPDTVGMNYGGSYMEEKHMPPPNMTTNERRVIVPADPTLWSTDHVRQWLEWAVKEYGLP
                     DVNILLFQNIDGKELCKMTKDDFQRLTPSYNADILLSHLHYLRETPLPHLTSDDVDKA
                     LQNSPRLMHARNTDLPYEPPRRSAWTGHGHPTPQSKAAQPSPSTVPKTEDQRPQLDPY
                     QILGPTSSRLANPGSGQIQLWQFLLELLSDSSNSSCITWEGTNGEFKMTDPDEVARRW
                     GERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPPESSLY
                     KYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSFFAAPNPYWNSPTGGIYPNTRLP
                     TSHMPSHLGTYY"
     repeat_region   291..298
                     /note="8 bp repeat sequence"
     repeat_region   509..516
                     /note="8 bp repeat sequence"
     polyA_signal    3143..3149
                     /note="putative"
BASE COUNT       915 a    770 c    768 g    713 t
ORIGIN       227 bp upstream of PvuII site; chromosome 21.


Alignment scores:
            Quality:  800.50              Length:    886
              Ratio:  1.598                 Gaps:     17
 Percent Similarity:  48.705   Percent Identity: 26.684

Alignment block:
 US-09-126-945B-2 x HUMERG2   ..
```

```
Align Seq 1/1  to: HUMERG2  from: 1  to: 3166

     45 ProSerProProAlaThrProGluGlnGlyLeuSerAlaPhe..... 58
        ||:::::||   |||      ::::::::::||||||:::
    272 CCGGACCCAGCAGCTCATATCAAGGAAGCCTTATCAGTTGTGAGTGAGGA 321

     59 .................................TyrLeuSerTyrPheA 64
                                        :::|||:::     :
    322 CCAGTCGTTGTTTGAGTGTGCCTACGGAACGCCACACCTGGCTAACACAG 371

     64 spMetLeu.................TyrProGluAspSerSerTrpAla 74
        :::||                  |||   :::   ||:::   :::
    372 AGATGACCGCGTCCTCCTCCAGCGACTATGGACAGACTTCCAAGATGAGC 421

     75 AlaLysAlaProGlyAlaSerSerArgGluGluProPro........... 87
        :::::::|||      :::      :::::|||||
    422 CCACGCGTCCCTCAGCAGGATTGGCTGTCTCAACCCCCAGCCAGGGTCAC 471

     88 .............................................GluGluP 96
                                                    :::::|
    472 CATCAAAATGGAATGTAACCCTAGCCAGGTGAATGGCTCAAGGAACTCTC 521

     90 roGluGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAsp 106
        ||::::::|||   |||      :::         :::|||||   |||
    522 CTGATGAATGCAGTGTGGCCAAAGGCGGGAAGATGGTGGGCAGCCCACAC 571

    107 LeuValPro.........GlyGlyLeuThrLeuGluGluHisSerLeuGl 120
        |||             ||:::          |||::|||
    572 ACCGTTGGGATGAACTACGGCAGCTACATGGAGGAGAAGCACATG..... 616

    120 aGlnValGlnSerMetValValGlyGlnValLeuLysAspIleGluThrA 137
                ::|||:::::::::|||
    617 .CCACCCCCAAACATGACCACGAACGAG..................... 643

    137 laCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSer 153
            :::   :::   :::   |||||||||      ||||||   :::
    644 .....CGCAGAGTTATCGTGCCAGCAGATCCTACGCTATGGAGTACAGAC 688

    154 AsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProPr 170
        :::|||:::::::|||||   |||:::      :::|||   ||||||
    689 CATGTGCGGCAGTGGCTGGACTGGGCGGTGAAAGAATATGGCCTTCCAGA 738

    170 oMetGly...LysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaM 186
        ::::::      |||||||:::::   |||||||||||||||   |
    739 CGTCAACATCTTGTTATTCCAGAACATCGATGGGAAGGAACTGTGCAAGA 788

    186 etSerGluGluGlnPheArgGlnArgSerPro...LeuGlyGlyAspVal 201
        |:::::::::::::|||:::::   :::|||      :::::|||:::
    789 TGACCAAGGACGACTTCCAGAGGCTCACCCCCAGCTACAACGCCGACATC 838

    202 LeuHisAlaHisLeu..................................... 206
        |||   :::||||||
    839 CTTCTCTCACATCTCCACTACCTCAGAGAGATCCCTCTTCCACATTTGAC 888

    207 .......AspIleTrpLysSer............................ 211
              ||:::   |||:::
    889 TTCAGATGATGTTGATAAAGCCTTACAAAACTCTCCACCGTTAATGCATG 938

    212 ..........................................AlaAlaTrp... 214
                                                :::|||||
    939 CTAGAAACACAGATTTACCATATGAGCCCCCCAGGAGATCAGCCTGGACC 988

    215 ............................MetLysGluArgThrSerProGl 222
                                    :::   ||||||||
    989 GGTCACGGCCACCCCACGCCCCAGTCGAAAGCTGCTCAACCATCTCCTC 1038

    222 yAlaIleHisTyrCysAlaSerThrSerGluGlu................ 233
        :::::::           :::|||:::::::::
   1039 CACAGTG.........CCCAAAACTGAAGACCAGCGTCCTCAGTTAGATC 1079

    234 ..............SerTrpThrAspSerGlnValAspSerSerCysSe 245
                     :::   |||:::|||:::::   :::      |||
   1080 CTTATCAGATTCCTGGACCAACAAGTAGCCGCCTTGCAAATCCAGGCAGT 1129

    246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPr 262
        ||||||   |||:::|||||||||||||   |||   ||||||:::
   1130 GGCCAG...ATCCAGCTTTGGCAGTTCCTCCTGGAG...CTCCTGTCGGA 1173

    262 oHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 279
           |||   :::      |||   |||   :::      ::::||   |
   1174 CAGCTCCAACTCCAGCTGCATCACCTGG...GAAGGCACCAACGGGCAGT 1220

    279 heLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLys 295
        ||||:::   |||      :::|||||||||   ||||||   ||||||
   1221 TCAAGATGACGGATCCCGACGAGGTGGCCCGGCGCTGGGGAGAGCGGAAG 1270

    296 AsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTy 312
        :::::::::   |||||||||||||||||||||||||||:::|||   ||
   1271 AGCAAACCCAACATGAACTACGATAAGCTCAGCCGCGCCCTCCGTTACTA 1320

    312 rTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVal 329
        ||||   |||:::|||:::   |||          :::::::|||   :::|
   1321 CTATGACAAGAACATCATCACCAAGGTCCAT...GGGAAGCGCTACGCCT 1367

    329 yrGlnPhe 336
        |||:::|
   1368 ACAAGTTC 1375

seq_name: qb_pr4:AF049703

seq_documentation_block:
LOCUS       AF049703      898 bp    mRNA            PRI       27-NOV-2000
DEFINITION  Homo sapiens E74 like factor 5 (ELF5) mRNA, complete cds.
ACCESSION   AF049703
VERSION     AF049703.1  GI:4928981
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   Zhou,J., Ng,A.Y., Tymms,M.J., Jermlin,L.S., Seth,A.K., Thomas,R.S.
            and Kola,I.
  TITLE     A novel transcription factor, ELF5, belongs to the ELF subfamily of
            ETS genes and maps to human chromosome 11p15: a region subject
            to LOH and rearrangement in human carcinoma cell lines
  JOURNAL   Oncogene 17 (21), 2719-2742 (1998)
  MEDLINE   99054671
   PUBMED   9840936
REFERENCE   2  (bases 1 to 898)
  AUTHORS   Zhou,J., Ng,A.Y.N., Tymms,M.J., Thomas,R.S. and Kola,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-1998) Molecular Genetics and Development Group,
            Institute of Reproduction and Development, Monash Medical Centre,
            246 Clayton Road, Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p13-p15"
     gene            1. .898
                     /gene "ELF5"
     CDS             110. .877
                     /gene "ELF5"
                     /note "Ets transcription factor"
                     /codon_start 1
                     /product "E74 like factor 5"
                     /protein_id "AAC79755.1"
                     /db_xref="GI:4928982"
                     /translation "MEESVIHSIEELFNASPGGILMSWTGEESNEEYYPAFFEGIACDS
                     YWTSVHPEYWTKKHVWEWLQFCQDQYKLDPNCISPCNFNISGLQLCSMLQEFFVFAAG
                     LCGEYLYFILQNIRIQGYSFENDAEESKALIKDYADSNCLKISGIKSQDCHSHSKISL
                     QSSEHEWEYEEELESHELNCSILEWEDEEGTIEVVEGLALALAEMWPGSSKNEGMIYEE
                     LSKALRYYYKTGILEKVDKRLVYKFGKNAHGWQEDKL"
BASE COUNT      246 a    221 c    219 g    212 t
```

```
ORIGIN


alignment scores:
            Quality:   300.00              Length:    262
              Ratio:   2.013                Gaps:      6
 Percent Similarity:   56.870   Percent Identity:  40.153

alignment block:
 US-09-126-945B-2 x AF049703   ..

 Align seq 1/1  to: AF049703  from: 1  to: 898

      102 AlaGlySerLeuAspLeuValProGlyGlyLeuThrLeu..... 114
          |||::|||||||||||||||:   ::::::||||||::
       71 GCCTCCTCTTTGGACCTAGCCACCGCTGCCCTCACGGTAATGTTGGACTC 120

      115 .GluGluHisSerLeuGlnGlnValGlnSerMetValValGlyGluVal 134
             ||||||          ::::::          ::::
      121 GGTGACACACAGCACCTTCCTGCCTAATGCATCCTTCTGCGATCCCCTGA 170

      135 euLys.....Asp 138
          :::::       :::
      171 TGTCGTGGACTGATCTGTTCAGCAATGAAGAGTACTACCCTGCCTTTGAG 220

      139 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIr 150
          :::|||||||||          ::::::   |||      ||
      221 CATCAGACAGCCTGTGACTCATACTGGACATCAGTCCACCCTGAATACTG 270

      150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrA 167
          |:::      ::::||   :::||||||   :::          ||||||
      271 GACTAAGCGCCATGTGTGGGAGTGGCTCCACTTCTGCTCGCACCAGTACA 320

      167 rqLeuProProMetGlyLysAlaPhe.....GlnGluLeuAlaGlyLys 181
          ::|||                ::|||          ::::::|||
      321 AGTTGGACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGGCCTG 370

      182 GluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGl 198
          :::|||||:::||||::::::|||::|||   :::   :::   |||
      371 CAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAGCTGGCCCTCTG 420

      198 yGlyAspValLeuHisAlaHisLeu..... 206
           ||::::   ||::::      |||
      421 CGGCGACTACCTGTACTTCATCTTCCAGAACATCCGCCACACAAGGTTACT 470

      207 ..........AspIleTrpLysSerAlaAlaTrpMetLysGluArgThr 219
                  |||      :::|||   |||   :::||::::   :::
      471 CCTTTTTTAATGACGCTGAAGAAAGCAAGGCCACCATCAAAGACTATGCT 520

      220 SerProGlyAlaIleHisTyrCysAlaSerThrSerGluGlnSerTrpTh 236
          :::   :::          |||   :::||||||      :::
      521 GATTCCAAC..........TGCTTGAAAACAAGTGGCATCAAAAGTCA 558

      236 rAspSerGluValAspSerSerCysSerGlyGluProIleHisLeuTrpC 253
          |||   :::      |||      |||   |||      |||||||||:
      559 AGACTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGG 608

      253 InPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 269
          :::|||:::::::::|||||||||:::|||::::::          ::::::
      609 AATTTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTG 658

      270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 286
          :::|||   ::::::|||:::::::::::::::::::::      |||
      659 GAATGGGAAGATAGGGAACAAGGAATTTTTCGGCTGGTTAAATCGGAAGC 708

      286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 303
          :::|||:::::::|||||   ||||||::::::      |||:::::|||
      709 CCTGGCAAAGATGTGGGACAAAGGAAGAAAAATGACAGAATGACATATG 758

      303 spLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArg 319
          ::|||||||||||||:::::::|||   ||||||||   ||||||::::::
      759 AAAAGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGGAATTTTGGAG 808

      320 LysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          :::   |||      :::|||||||||||||:::|||
      809 CGGGTTGAC.....CGAAGTTAGTGTACAAATTT 838

 Seq_name: qb_pat1:AR103486

 Seq_documentation block:
 LOCUS       AR103486     2280 bp    DNA              PAT       14-FEB-2001
 DEFINITION  Sequence 8 from patent US 6087485.
 ACCESSION   AR103486
 VERSION     AR103486.1  GI:12815074
 KEYWORDS    .
 SOURCE      Unknown.
   ORGANISM  Unknown.
             Unclassified.
 REFERENCE   1  (bases 1 to 2280)
   AUTHORS   Brooks Wilson,A.R., Buckler,A., Cardon,L., Carey,A.H., Galvin,M.,
             Miller,A. and North,M.
   TITLE     Asthma related genes
   JOURNAL   Patent: US 6087485-A 8 11-JUL-2000;
 FEATURES             Location/Qualifiers
      source          1..2280
                      /organism="unknown"
 BASE COUNT      676 a    481 c    472 g    651 t
 ORIGIN


 alignment scores:
             Quality:   300.00              Length:    262
               Ratio:   2.013                Gaps:      6
  Percent Similarity:   56.870   Percent Identity:  40.153

 alignment block:
  US-09-126-945B-2 x AR103486   ..

  Align seq 1/1  to: AR103486  from: 1  to: 2280

      102 AlaGlySerLeuAspLeuValProGlyGlyLeuThrLeu..... 114
          |||::|||||||||||||||:   ::::::||||||::
       82 GCCTCCTCTTTGGACCTAGCCACCGCTGCCCTCACGGTAATGTTGGACTC 81

      115 .GluGluHisSerLeuGlnGlnValGlnSerMetValValGlyGluVal 134
             ||||||          ::::::          ::::
       82 GGTGACACACAGCACCTTCCTGCCTAATGCATCCTTCTGCGATCCCCTGA 131

      135 euLys.....Asp 138
          :::::       :::
      132 TGTCGTGGACTGATCTGTTCAGCAATGAAGAGTACTACCCTGCCTTTGAG 181

      139 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIr 150
          :::|||||||||          ::::::   |||      ||
      182 CATCAGACAGCCTGTGACTCATACTGGACATCAGTCCACCCTGAATACTG 231

      150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrA 167
          |:::      ::::||   :::||||||   :::          ||||||
      232 GACTAAGCGCCATGTGTGGGAGTGGCTCCACTTCTGCTCGCACCAGTACA 281

      167 rqLeuProProMetGlyLysAlaPhe.....GlnGluLeuAlaGlyLys 181
          ::|||                ::|||          ::::::|||
      282 AGTTGGACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGGCCTG 331

      182 GluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGl 198
          :::|||||:::||||::::::|||::|||   :::   :::   |||
      332 CAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAGCTGGCCCTCTG 381

      198 yGlyAspValLeuHisAlaHisLeu..... 206
           ||::::   ||::::      |||
      382 CGGCGACTACCTGTACTTCATCTTCCAGAACATCCGCCACACAAGGTTACT 431

      207 ..........AspIleTrpLysSerAlaAlaTrpMetLysGluArgThr 219
                  |||      :::|||   |||   :::||::::   :::
      432 CCTTTTTTAATGACGCTGAAGAAAGCAAGGCCACCATCAAAGACTATGCT 481
```

```
      220 SerProGlyAlaIleHisTyrCysAlaSerThrSerGluGlaSerIrpIh 269
          :::   :::          |||   :::|||||      :::
      482 GATCCAAC.....TGCTTGAAAACAAGTGGCATCAAAAGTCA 519

      245 rAspSerGluValAspSerSerCysSerGlyGlaProIleHisLeuTrpC 294
          |||   :::      |||     |||   |||     |||||||||
      520 AGACTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGG 569

      253 lnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 302
          ::|||:::::::::::|||||:|||:::|||||:::        ::::::
      570 AATTTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTG 619

      270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaCl 319
          :::|||   :::::::|||:::|||||||||::::::      |||
      620 GAATCGGAAGATAGGGAACAAGGAATTTTTCGGGTGGTTAAATCGGAAGC 669

      286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 335
          :::|||::::::|||||   ||||||:::::      ||||:|||||
      670 CCTGGCAAAGATGTGGGGACAAAGGAAGAAAAATGACAGAATGACCTATG 719

      304 spLysLeuSerArgSerIleArgGlaTyrTyrLysLysGlyIleIleAr; 353
          ::|||||||||||||:::::|||   |||||||||   ||||||:::::
      720 AAAAGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGGAATTTTGGAG 769

      320 LysProAspIleSerGlnArgLeuValTyrGlnPhe 351
          :::   |||      :::|||||||||||:|||
      770 CGGCTTGAC.....CGAAGGTTAGTGTACAAATTT 799

Seq_name: qb_pat1:AR103487

Seq_documentation_block:
LOCUS       AR103487     2498 bp    DNA              PAT       14-FEB-2001
DEFINITION  Sequence 10 from patent US 6087485.
ACCESSION   AR103487
VERSION     AR103487.1  GI:12815075
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2498)
  AUTHORS   Brooks Wilson,A.R., Buckler,A., Cardon,L., Carey,A.H., Galvin,M.,
            Miller,A. and North,M.
  TITLE     Asthma related genes
  JOURNAL   Patent: US 6087485-A 10 11-JUL-2000;
FEATURES             Location/Qualifiers
     source          1..2498
                     /organism="unknown"
BASE COUNT      706 a    553 c    554 g    685 t
ORIGIN


Alignment_scores:
            Quality:  299.50              Length:     450
              Ratio:  1.576                 Gaps:      13
  Percent Similarity:  54.286   Percent Identity: 29.429

Alignment_block:
 US-09-126-945B-2 x AR103487   ..

 Align seq 1/1  to: AR103487  from: 1  to: 2498

        8 SerAlaSerProGlyLeu.SerSerValSerProSerHisLeuLeuLeuP 19
          :::|||||   ||||||   :::::::::   ||||||      :::|
       85 GCGGCCTCTGGAGGCCTGCCATGGCACCCAGAGCAGCCAT......TTTG 128

       19 roProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAla 45
          ||||||:::::                     |||      ||||||
      129 CTCCCAGTTCT.....................GGGGCTTTGGAAGGAGCT 157

       46 ValGly.LeuGluArgArg......AspTrpSerProSerProProAlaI 50
             |||   |||||||:::      :::   |||   :|||||||||:::
      158 TGCGGATGAGGAGAGGGAGCCTCCGCAGGGCTCTGGCTCCCCTCCAGGGG 207

       50 hrProGluGlaGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeu 66
          ::                ::::||                     |||
      208 CCGAGGCCGCACAC'AAAGCCGCT.....................CTG 253

       67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 83
          :::|||:::   :::  |||   :::        |||
      254 TGGCCCAATTACACCTACTGGATAGGATTGTTGAGGGGA..... 272

       83 qGluGlnProProGlnGluProGlaGlnCysProValIleAspSerGlnA 100
                |||||||:::   |||          :::   ||||||||||
      273 .........CCTGAGAAACTTGAGACCACAAGAACGCGTAGCGCCTCGG 312

      100 laProAlaGlySerLeuAspLeuValPro:GlyGlyLeuThrLeuGlu... 115
          ||      |||          ||||||    :::|||   |||:::
      313 CTGGCTGAGGGTGCTGAGCCCTCGTGTTGTTCTCTCCAGCTTTGCCCG 362

      116 ......GlaHisSerLeuGluGlnValGluSerMetValValGlyGluVa 130
                :::|||||!   ||||||||||   ::::::
      363 TGCCTCAGCCACTCTTCACGTTCCATCTGTGCTCTG.............. 398

      130 IleuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspP 147
                              ||||||                 ::::::  |
      399 ..........TGCTGACCCGCCTGTGACTCATACTGGACATCAGTCCACC 448

      147 roMetAspIrpSerProSerAsnValGluLysIrpLeuLeuIrpIhrGlu 163
          ||      ||:::::     :::|:|    ::::||||||   :::
      449 CTGAATACTGGACTAAGCGCCATGTGTGGGAGTGGCTCCAGTTCTGCTGC 498

      164 HisGlnTyrArgLeuProProMetGlyLysAlaPhe......GlnGluLe 178
             ||||||:::|||                  :::|||        :::::
      499 GACCAGTACAAGTTGGACACCAATTGCATCTCCTTCTGCAACTTCAACAT 548

      178 uAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgS 195
          :::|||  :::|||||||:::|||::::::::|||:::|||   :::   :
      549 CAGTGGCCTGCAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAG 598

      195 erProLeuGlyGlyAspValLeuHisAlaHisLeu.............. 206
          ::   |||   ||:::   |||:::      |||
      599 CTGGCCTCTGCGGCGAGTACCTGTACTTCATCCTCCAGAACATCCGCACA 648

      207 ....................AspIleIrpLysSerAlaAlaIrpMetLy 216
                                |||      :::|||   |||   :::||
      649 CAAGGTTACTCCTTTTTAATGACCCTGAAGAAAGCAAGGCCACCATCAA 698

      216 sGluArgIhrSerProGlyAlaIleHisTyrCysAlaSerIhrSerGluG 233
          |:::   ::::::   :::          |||   :::|||||
      699 AGACTATGCTGATCCAAC.....TGCTTGAAAACAAGTGGCA 726

      233 luSerIrpIhrAspSerGluValAspSerSerCysSerGlyGlnProIle 249
          :::      |||   :::      |||      |||   |||
      727 TCAAAAGTCAAGACTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCT 776

      250 HisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGl 266
          ||||||||||:::|||:::::::::|:|||||||||:::|||:::::::
      777 CATCTATGGGAATTTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTG 826

      266 yArgPheIleArgIrpLeuAsnLysGluLysGlyIlePheLysIleGluA 283
          :::::::::::|'   :::::|||:::::::::::::::::::::::
      827 TGGCATTCTGGAATGGGAAGATAGGGAACAAGGAATTTTTCGGGTGGTTA 876

      283 spSerAlaGlnValAlaArgLeuIrpGlyIleArgLysAsnArgProAla 299
          |||      :::|||:::::::|||||   ||||||:::::
      877 AATCGGAAGCCTGGCAAAGATGTGGGGACAAAGGAAGAAAAATGACAGA 926

      300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 316
          |||:::|||:::|||||||||||||:::::::::   |||||||||   ||
      927 ATGACGTATGAAAAGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGG 976

      316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 351
          ||||:::::::::   |||      :::|||||||||||:::|||
      977 AATTTTGGAGCGGGTTGAC.....CGAAGGTTAGTGTACAAATTT 1016

Seq_name: qb_pr10:HUMEIS2A
```

```
seq_documentation_block:
LOCUS        HUMETS2A     2269 bp    mRNA            PRI       08 NOV 1994
DEFINITION   Human erythroblastosis virus oncogene homolog 2 (ets 2) mRNA,
             complete cds.
ACCESSION    J04102
VERSION      J04102.1  GI:182272
KEYWORDS     ets 2 gene.
SOURCE       Human K562 cell line, cDNA to mRNA.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 2269)
  AUTHORS    Watson,D.K., McWilliams,M.J., Lapis,P., Lautenberger,J.A.,
             Schweinfest,C.W. and Papas,T.S.
  TITLE      Mammalian ets-1 and ets-2 genes encode highly conserved proteins
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 85 (21), 7862-7866 (1988)
  MEDLINE    89042086
COMMENT      Draft entry and computer-readable sequence [1] kindly submitted by
             J.Lautenberger 07-OCT-1988.
FEATURES             Location/Qualifiers
     source          1..2269
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="21q22.3"
     gene            292..1701
                     /gene="ETS2"
     CDS             292..1701
                     /gene="ETS2"
                     /note="ets2 protein"
                     /codon_start=1
                     /db_xref="GDB:G00-119-888"
                     /protein_id="AAA52412.1"
                     /db_xref="GI:182273"
                     /translation="MNDFCLKNMIHAVPVANSYKPTLEKHQPAPLPTPIHSTPAVFPSLN
                     EEQTHQEVPTGLDSLTSHDSANCELPLLTPCSKAVMSQALKATPSGFKKEQRRLGTPKN
                     PWLWSEQQVCQWLLWATNEFSLVNVNHQRFGMNGQMLCNLGKERFLELAPDFVGDILW
                     EHLEQMIKENQEKTEDQYEENSHLTSVPHWFNSNTIGFGTEQAPYGMQTQNYPKGGLL
                     DSMCPASTPSVLSSFQEFQMFPKSFISSVSVTYCSVSQDFPGSNLNLLTNNSGTPKDH
                     DSPENGADSFESSDSLLQSWNSQSSLLDVQRVPSFESFEDDCSQSLCLNKPTMSFKDY
                     IQERSDPVEQCKPVTPAAVLACPTCQSGPTQLWQFLLELLSDKSCQSFISWTGDGWEFK
                     LADPDEVARRWGKRKNKPKMNYEKLSRGLRYYYDKNIIHKTSGKRYVYRFVCDLQNLL
                     GFTPEELHAILGVQPDTED"
BASE COUNT       550 a    643 c    584 g    492 t
ORIGIN       Chromosome 21q22.3.


alignment scores:
            Quality:  299.00              Length:    552
              Ratio:  1.384                 Gaps:     18
 Percent Similarity:  39.130   Percent Identity: 22.464

alignment block:
 US-09-126-945B-2 x HUMETS2A  ..

 Align seq 1/1  to: HUMETS2A  from: 1  to: 2269

        8 SerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuPr 19
          |||:::  |||      |||:::|||  ||||||      ||||||||
        6 TCCTCCCCTCCCCTCCACTCGGCCGTCCCTCCTTCCTCCTCCCTCCTCCC 55

       19 oProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaV 36
          ||||      :::  :::         |||:::||||||  |||:::
       56 TCCTCCTCCCGCTCCTGAA.....GAGCGGGCCGCGTGGGGGACGG 96

       46 alGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu 52
          ::  |||                          ||||||  |||  |||
       97 CCCGGTTA.....................CTTCCTCCAGAGACTGACGAG 125

       54 GluGlyLeuSerAlaPheTyrLeuSerTyrPhe.....AspMetLeuTy 67
             |||:::::             |||             :::
      126 TGCGGTGTCGCTCCAGCTCAGACGTCCCGGAGCCGCCCGGCCAGCGTCCG 175

       67 rProGluAspSerSer.TrpAlaAlaLysAlaPro..............  78
            |||      ||||||  |||   |||   :::|||
      176 GCCTCCCTGATCGTTCTCGGCCGGCGCCCPCGCCCTCGCCCCGCCGCCAC 225

       79 ...GlyAlaSerSerArgGluGluProProGluGluProGluGlnCysPr  94
             :::|||:::::   |||:::||||||:::::|||:::      ||
      226 CGAGCAGCCCCCGCCGCCCAGCAGCCACCGTCCCGACCAAGCCCCGCCCC 275

       94 oVal.................................IleAspSerG 99
          |:::                                 :::||||::
      276 TGCCCGCAGCGGCAGGATGAATGATTTCGGAATCAAGAATATGGACCAGG 325

       99 lnAlaProAlaGly..................................  103
            ||||||:::::
      326 TAGCCCCTGTGGCTAACAGTTACAGAGGGACACTCAAGCGCCAGCCAGCC 375

      103 .....................................................  103

      376 TTTCACACCTTTCATGGGTCCCTGTTTGCTGTTTTTCCTTCTCTAAATGA 425

      104 .......SerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisS 118
                :::|||:::   ||||||   ||||||         :::|||:
      426 AGAGCAAACACTGCAAGAAGTGCCAACAGGCTTGGATTCCATTTCTCATG 475

      118 er.....................LeuGluGlnValGlnSerMetVal 126
          ::                  |||         ::::::   |||
      476 ACTCCGCCAACTGTGAATTGCCTTTGTTAACCCCGTGCAGCAAGGCTGTG 525

      127 ValGlyGluValLeuLysAspIleGluThrAlaCysLys..........  139
          :::::::::::::||||||         ::::::   |||
      526 ATGAGTCAAGCCTTAAAAGCTACCTTCAGTGGCTTCAAAAAGGAACAGCG 575

      140 _LeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValG 156
          |||:::|||     :::|||     ||||||   ::::::|||
      576 GCGCCTGGGCATTCCAAAGAACCCCTGCCTCTGCACTCACCAACAGGTAT 625

      156 lnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGly 172
          :::|||||||||||||:::   :::::::::   |||       ::::::
      626 GCCAGTGGCTTCTCTGGGCCACCAATGAGTTCAGTCTGGTGAACGTGAAT 675

      173 LysAlaPheGlyGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGl 189
                     :::   ||:::   ||||||   ::::::::::||
      676 CTGCAGAGGTTCGGCATGAATGGCCAGATGCTGTGTAACCTTGGCAAGGA 725

      189 uGlnPheArgGluArgSerPro...LeuGlyGlyAspValLeuHisAlaH 205
          |:::|||   :::   :::|||   :::   ||||||::|||      |
      726 ACGCTTTCTGGAGCTGGCACCTGACTTTGTGGGTGACATTCTCTGGGAAC 775

      205 isLeuAsp.............................................  207
          |||||:::
      776 ATCTGCAGCAAATGATCAAACAAAACCAACAAAAGACACAAGATCAATAT 825

      208 ..............IleTrpLysSerAlaAlaTrpMet............  215
                       :::   :::        |||:::
      826 GAAGAAAATTCACACCTCACCTCCGTTCCTCATTGGATTAACAGCAATAC 875

      215 .....................................................  215

      876 ATTAGGTTTTGGCACAGAGCAGGCGCCCTATGGAATGCAGACACAGAATT 925

      215 .....................................................  215

      926 ACCCCAAAGGCGGCCTCCTGGACAGCATGTGTCCGGCCTCCACACCCAGC 975

      216 ..................................LysGluArgThrSe 220
                                           |||:::|||   ||
      976 GTACTCAGCTCTGAGCAGGAGTTTCAGATGTTCCCCAAGTCTCGGCTCAG 1025

      220 rProGlyAlaIleHisTyrCysAlaSerThrSerGlu.............  232
          |     ::::::  ||||||:::  :::::::::
     1026 CTCCGTCAGCGTCACCTACTGCTCTGTCAGTCAGGACTTCCCAGGCAGCA 1075

      232 .....................................................  232
```

```
    1076 ACTTGAATTTGCTCACCAACAATTCTGGGACTCCCAAAGACCACGACTCC 1125

     233 .....GluSe 234

    1126 CCTGAGAACGGTGCGGACAGCTTCGACAGCTCACACTCCCTCCTCCACTC 1175

     234 rTrpThrAsp.....S 238

    1176 CTGGAACAGCCAGTCGTCCTTGCTGGATGTGCAACGGGTTCCTTCCTTCG 1225

     238 erGluValAspSerSerCysSer..... 245

    1226 AGAGCTTCGAACATGACTGCAGCCAGTCTCTCTGCCTCAATAAGCCAACC 1275

     245 ..... 245

    1276 ATGTCTTCAAGGATTACATCCAAGACAGGAGTGACCCAGTGGAGCAAGG 1325

     246 .....GlyGlnP 248

    1326 CAAACCAGTTATACCTGCAGCTGTGCTGGCCGGCTTCACAGGAAGTGGAC 1375

     249 roIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSer 264

    1376 CTATTCAGCTGTGGCAGTTTCTCCTGGAG...CTGCTATCAGACAAATCC 1422

     265 TyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIl 281

    1423 TGCCAGTCATTCATCAGCTGGACTGGA...GACGGATGGGAGTTTAAGCT 1469

     281 eGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgP 298

    1470 CGCCGACCCCGATGAGGTGCCCGCCGGTGGGGAAAGAGCAAAAATAAGC 1519

     298 roAlaMetAsnTyrAspLysLeuSerArgSerIleArgIleTyrTyrLys 314

    1520 CCAACATCAACTACGACAAGCTGAGCCGGGGCTTACGCTACTATTACGAC 1569

     315 LysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPh 331

    1570 AAGAACATCATCCACAAGACGTCG...GGGAAGCGCTACGTGTACCGCTT 1619

     331 eVal 332

    1617 CGTG 1620
```

seq_name: qb_pat1:AR103485

```
seq_documentation_block:
LOCUS       AR103485     2428 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6087485.
ACCESSION   AR103485
VERSION     AR103485.1  GI:12815073
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2428)
  AUTHORS   Brooks-Wilson,A.R., Buckler,A., Cardon,L., Carey,A.H., Galvin,M.,
            Miller,A. and North,M.
  TITLE     Asthma related genes
  JOURNAL   Patent: US 6087485-A 6 11-JUL-2000;
FEATURES             Location/Qualifiers
     source          1..2428
                     /organism="unknown"
BASE COUNT      705 a    519 c    519 g    685 t
ORIGIN
```

alignment_scores:
Quality: 299.00 Length: 211
Ratio: 2.318 Gaps: 4

Percent Similarity: 61.137 Percent Identity: 33.175

alignment_block:
US-09-126-945B-2 x AR103485 ..

Align seq 1/1 to: AR103485 from: 1 to: 2428

```
     135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIrpSe 151

     333 CAGACAGCCTGTGACTCATACTGGACATCAGTCCACCCTGAATACTGGAC 382

     151 rProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 168

     383 TAAGCGCCATGTGTGGGACTGGCTCCAGTTCTGCTGCGACCAGTACAAGT 432

     168 euProProMetGlyLysAlaPhe......GlnGlnLeuAlaGlyLysGln 182

     433 TGCACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGGCCTGCAG 482

     183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 199

     483 CTGTGCAGCATGACACAGGAGGAGTTCCGTCGAGGCAGCTGGCCTCTGCGG 532

     199 yAspValLeuHisAlaHisLeu..... 206

     533 CGAGTACCTGTACTTCATCCTCCAGAACATCCGCACACAAGGTTACTCCT 582

     207 ........AspIleTrpLysSerAlaAlaTrpMetLysGluArgThrTer 220

     583 TTTTTAATGACGCTGAAGAAAGCAAGGCCACCATCAAACACTATGCTGAT 632

     221 ProGlyAlaIleHisTyrCysAlaSerThrSerGluGlnSerTrpThrAs 237

     633 TCCAAC  TGCTTCAAAACAAGTGGCATCAAAAGTCAAGA 670

     237 pSerGlnValAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnP 254

     671 CTCTCACAGTCATAGTACAACAAGCCTCCAAAGTTCTCATCTATGGCAAT 720

     254 heLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArg 270

     721 TTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTGGAA 770

     271 TrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVa 287

     771 TGGGAACATACGGAACAAGCAATTTTTCGGGTGGTTAAATCGGAAGCCCT 820

     287 lAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspL 304

     821 GGCAAAGATGTGGGGACAAAGGAAGAAAAATGACAGAATGACGTATGAAA 870

     304 ysLeuSerArgSerIleArgIleTyrTyrLysLysGlyIleIleArgLys 320

     871 AGTTGAGCAGAGCCCTGAGATACTACTATAAAACAGGAATTTTGGAGCGG 920

     321 ProAspIleSerGlnArgLeuValTyrGlnPhe 331

     921 GTTGAC......CGAAGGTTAGTGTACAAATTT 947
```

seq_name: qb_in3:CELF22A3

```
seq_documentation_block:
LOCUS       CELF22A3     29997 bp    DNA             INV       08-DEC 1995
DEFINITION  Caenorhabditis elegans cosmid F22A3.
ACCESSION   U41547
VERSION     U41547.1  GI:1109905
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 29997)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
```

```
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 29997)
  AUTHORS   Hallsworth,K
  TITLE     The sequence of C. elegans cosmid F22A3
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 29997)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1995) Robert Waterston
COMMENT                 Submitted by:
                        Genome Sequencing Center
                        Department of Genetics, Washington University,
                        St. Louis, MO  63110, USA, and
                        Sanger Centre, Hinxton Hall
                        Cambridge CB10 1RQ, England
                        e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
                        NEIGHBORING COSMID INFORMATION:
            5' cosmid is C15B12, 299 kp overlap, 3' cosmid is C23H3, 299 kp
            overlap. Actual start of this cosmid is at base position 39847 of
            CELC15B12; actual end is at 4499 to CELT14E8.

            NOTES:
            Coding sequences below are predicted from computer analysis, using
            the program GeneFinder(P. Green and L. Hillier, ms in preparation).
FEATURES                Location/Qualifiers
     source             1..29997
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /map="X"
     gene               327..2922
                        /gene="F22A3.4"
     CDS                join(327..408,726..981,989..1129,1177..1278,1317..1401,
                        1984..2158,2448..2673,2722..2922)
                        /gene="F22A3.4"
                        /note="Similar to glutamate receptor."
                        /codon_start=1
                        /evidence=not_experimental
                        /protein_id="AAA83192.1"
                        /db_xref="GI:1109906"
                        /translation="MEYIRSVIAFLIYPPQFPPTGIMGRSPLEYIEQQLALVIEPQW
                        EVWVIIAAVILISGVIFLVMTNFIRKVYIEMTVIPFEIIWVFGSIFVQQGLPEQPKW
                        SCRVLVALWWLASIILSATFTGSLVALFAVDEINVFLQNLPQLVEVSDQIELIVMGL
                        NFPYAYGCKVILPNGAILIVYLSIILVKNSIYSIYLIQKISLMVLDSILQEWIAGAYS
                        YVAMQKINFCNSFIGPKSYLDLKRAQGAFWVFLGGAGLGLALFVGFEIFKEFREKMT
                        KHTKTFTVPPEVATVSANVAENSTSLDIPDEVLPSAPLPSED@NWSLRRRNILNLSLD
                        LSNIAN"
     gene               3998..5870
                        /gene="F22A3.2"
     CDS                join(3998..4139,4193..4329,4372..4471,5443..5444,
                        5480..5689,5733..5870)
                        /gene="F22A3.2"
                        /note="coded for by C. elegans cDNA CEESO85F; similar to
                        C. elegans proteins C40H1.5, R13A5.3, F40F12.1, K03H1.3,
                        K03H1.4, K03H1.6, T07C4.5, F40F12.1, and R13A5.6"
                        /codon_start=1
                        /protein_id="AAA83193.1"
                        /db_xref="GI:1109907"
                        /translation="MFTIVITFLIGCAFAFPTQSIGVFGPLMGSSPAAGVNLPLPI
                        EDNGPDPDIVLDQKTTDDDGNFLLSGSSMELIPTDPELRIFHDCNDQGSPGQREWVIR
                        IPAKYIINGPEVKETMDLGVLNLEVEMISKLLIIGIGVLSVVLAASESHSGRQNYRVK
                        GAFRGGNVFVKNVQVKLIDDDFGSDPTDDDLGSGYINANGEFELSGSTIELIIIGHLK
                        IYHDGDLSINDGQRRWKFELPNNYIYSDDIPKIFDIGIWNLEGILPGESRDGNH"
     gene               9441..14972
                        /gene="F22A3.1"
     CDS                join(9441..9603,11480..11427,12165..12248,12290..12438,
                        12853..12936,12988..13027,13505..13651,13698..13889,
                        14021..14106,14214..14289,14342..14443,14491..14850,
                        14905..14972)
                        /gene="F22A3.1"
                        /note="coded for by C. elegans cDNA yk45a1.5; coded for by
                        C. elegans cDNA yk45a1.4; Similar to DNA binding protein.
                        EIS domain."
                        /codon_start=1
                        /protein_id="AAA83194.1"
                        /db_xref="GI:1109908"
                        /translation="MSSPLPSIDEIGSIGMSIWSGIPKEQASIGIESPESVPGLKRVG
                        AVAARRAPPPEPPPPPPTPTLLFKLSETQFHGLNQFAFPENNLEYKYGSGKSIDSSS
                        MNGIGSVGHRWNSISPEPHSGIESIASIPEVKSEEPEDDDEEGIDQVNNVKPHFMDMP
                        CNLPTQPIEYNKRFSKDADHSTFVKNEIEENILNENVNPETAQINGLDIQQIDIYRDL
                        IGRILIQDISIIGAKIGLPNDFYLWSSEHGARWINEMGMQFNLQPPRNGSTIGIDLIG
                        MSQKDFEMILPAGGDILHAQLQVWKTAPENYHPPVTVQSSGMIAAENNMQSKINWLAS
                        INNQINNMAAAENPNHPTPNGNGGYPNMGMSSTPQQGTVLPSPSNSDTSGNGSSQDMN
                        DIDIDLHMNNSNFGPSNFPHNQGYMNSPIDAMGNGSEGDDDERAYTRHQGIVHLWQFT
                        RELLDQPKQYSACVRWVDRIGLPKAIESSLLLAGYWGQKKNKSGMNYDKLSKSLIQYY
                        KKGLIQKPEKKQKLVYKFLPPYNL"
     gene               complement(17421..17630)
                        /gene="F22A3.4"
     CDS                complement(17421..17630)
                        /gene="F22A3.4"
                        /note="similar to human E2A-Pbx1 fusion protein and human
                        homeotic protein PBX2"
                        /codon_start=1
                        /evidence=not_experimental
                        /protein_id="AAA83195.1"
                        /db_xref="GI:1109909"
                        /translation="MQEDERERAASMANEAQAIQNSLNSSTVASSSNMSTSGLINIPL
                        VNPTMQSMVIPAVQPGLLGNPNGQL"
     gene               complement(19668..20721)
                        /gene="F22A3.5"
     CDS                complement(join(19668..19754,19980..20115,20597..20721))
                        /gene="F22A3.5"
                        /note="Similar to homeotic protein PBX2."
                        /codon_start=1
                        /evidence=not_experimental
                        /protein_id="AAA83196.1"
                        /db_xref="GI:1109910"
                        /translation="MEQPESNIMQSTSEKEEHSENNIPGAAIEEIVERPEIPQQQEE
                        EENEQENIIEEIQNWERGRQNEYIEQEAAIARQNIEEIQQNEIEKEHIEIIKS
                        GVINLEIDDLPQL"
     gene               complement(26607..29835)
                        /gene="F22A3.6"
     CDS                complement(join(26607..26705,27459..28198..28314,
                        29357..29469,29802..29835))
                        /gene="F22A3.6"
                        /codon_start=1
                        /evidence=not_experimental
                        /protein_id="AAA83197.1"
                        /db_xref="GI:1109911"
                        /translation="MFVKSLLLLSVATAYVSADCEHCTCMREGGCKPDGCHMGVGFS
                        GGYYQIEDCYYEIGGPIEKACEIIEAAWERGADENCATICVENYYNEYESGCNGLG
                        MGACGIHCRNHNGCPRGCHNANILAYWNGVKSGGGS"
BASE COUNT      9488 a   5249 c   4917 g  10443 t
ORIGIN


alignment scores:
                Quality:  298.50              Length:    472
                  Ratio:  1.745                 Gaps:     10
 Percent Similarity:  86.441   Percent Identity: 19.703

alignment block:
 US-09-126-945B-2 x CELF22A3    ..
```

```
Align seq 1/1  to: CELF22A3  from: 1  to: 29997

    114 LeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
        ||||::   :::::::::::      :::::::::::::::   :::::
  13555 CTTGACACCCAGCAAATTGACATCTACAGGGACCTGATTCTCCGTCATCT 13604

    130 IleLysAspIleGluThrAlaCysLysLeuLeuAsn..............  142
        ::::::|||||::::|||::|||      ||||::
  13605 GATTCAAGACATCAGTACAACTTGTGCCAAATTGGGACTTCCAAATGGTA 13654

    143 ................................IleThrAla.AspProMe  148
                                        ::::::::: |||
  13655 CATTACTTATCCACAATACCTCATTTATATTTTGGTCTCTAGATTTCTA 13704

    148 tAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisG 165
        ||||||   ::::::      :::||||::                   |
  13705 CCTGTGGTCATCCGAGCATGGAGCACGATGGATCAACGAGATGTGCATGC 13754

    165 hTyrArgLeuProProMetGlyLysAlaPheGluGluLeuAlaGlyLys 181
        ||::::::|||   |||      ::::::      :::::::::|||
  13755 AATTCAAICTTCAGCCGCCTAGAAACTGT......TCAATCACTGGTATT 13798

    182 GluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGl 198
        ::::||   :::|||||||:::::::::|||::::::      |||   ||
  13799 GATTTGCTCGGCATGAGTCAAAAAGATTTCGAAATGATTTTGCCGGCGGG 13848

    198 yGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAla.... 213
        |||||||:::|||||||||:::|||::::::||||||:::::::::
  13849 AGGCGACAGCGTTGCATGCCCAACTCCAAGTGTGCAAGACCGGTACGTCAG 13898

    213 ................................................... 213

  13899 ATTACGTGAAAGGTGATGCGATTAGCATGACTATAAATCACGCCCTTTTT 13948

    214 ...............................IrpMet.LysGluArgThrS 220
                                        :::   ||:::   :
  13949 CTGATTGTCATCATGACGGTTATCAGTGACATGCTCACACGAAAACGACA 13998

    220 erPro............................................... 221
        ::|||
  13999 CACCAATCCAATTCAATTTCAGCATTTGAAAACTATCATCCACCCGTGAC 14048

    221 ................................................... 221

  14049 CGTTCAGAGCTCCGGAATGACCGCTGCCGAAAATAACATGCAGTCAAAAA 14098

    222 ..................................................GlyA 223
                                                         ||||
  14099 CTAATTGCGTATGTAAATGTTGATTATGTCATTCCTCAGAATTTTCCGGC 14148

    223 laIleHisTyrCys..................................... 227
        ||    :::|||
  14149 CGCAGTCTTTTTGTACCCCTATTTGCTCAATGCGTTTTTTTCATAAAGCC 14198

    228 ..................AlaSerThrSerGluGluSer............ 234
                         ||||||||||:::::::::::
  14199 TTCTCGTATTTTTAGCTAGCCAGCACCAACAATCAAACAAACAACATGGC 14248

    234 ................................................... 234

  14249 AGCTGCTGAGAACCCGAATCATCGATTTTTCAATGGAAACGCTTGGTTTC 14298

    234 ................................................... 234

  14299 CAACCGCAACCCTTTGATGTGAATAAGTAAATTGTATTTATACGACGCTA 14348

    234 ................................................... 234

  14349 TCCAAACATGTCGATGAGCAGCTTCTTTCAACAGGGAACTGTTTTGCCAT 14398

    234 ................................................... 234

  14399 CGCCAAGCAACAGCGATACCAGCAGCAATGGGAGCAGTCAAGACAGTATG 14448

    234 ................................................... 234

  14449 ACACGGGCTTATTAGTACTGTAACTAATTCAGAGTTTTTTAGTGAATGAC 14498

    234 ................................................... 234

  14499 GATGATATTGATCTTCACATGAACAATTCGAACTGTGGTTTCTCAAACTT 14548

    235 ................TrpThrAspSerGluValAspSerSerCysSerG 246
                        :::   ::::|||   :::|||:::   |||:::|
  14549 CTTTCACAACCAGGGTTACATGAACTCCCCTATTGACGCAATGTGCAACG 14598

    246 lyGln.....................................Pro 248
        |||::
  14599 GAAGCGAGGGAGATGACGATGAGCGTGCCTACACTCGTCATCAAGGAACT 14648

    249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
        :::|||||||||||||||:::::|||||||||   :::|||   :::||
  14649 GTTCATCTGTGGCAGTTTATTCGTGAGCTTTTGGATCAACCAAAACAGTA 14698

    265 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleG 282
        |:::      :::||||||:::::::::::::::::|||   ||||||||||
  14699 CAGTGCATGTGTTCGATGGGTGCATAGACATCAGGGAACTTTTAAAATCG 14748

    282 luAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgPro 298
        ||:::|||      :::|||||||   ||||||   ||||||||||||
  14749 AGTCTTCCCTCCTGCTGGCTAGATATTGGGGACAACGAAAAAACCGTTCA 14798

    299 AlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLy 315
           ||||||||||||||||||||||||||:::|||||||||||||||||
  14799 CAGATGAATTATGACAAGTTGTCCAGAAGTCTTCGACAGTACTATAAGAA 14848

    315 s................................................... 315
        |
  14849 AGGTAAAAAGAATAGGAGGAAGTGTAAAACAAAAACGTTACAAATCAATA 14898

    316 .....GlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGln 330
             ||||||||||:::|||||:::   :::|||||||||||||||:::
  14899 TTCCAGGAATCATCCAAAAGCCGGAAAAGAAGCAGCGTTTGGTGTACAAA 14948

    331 PheValHisPro 334
        |||:::   |||
  14949 TTCCTGCCTCCT 14960

seq_name: qb_ov:XLAJ4125

seq_documentation_block:
LOCUS       XLAJ4125     1798 bp    mRNA            VRT       21-JAN-2000
DEFINITION  Xenopus laevis erg gene (erg_A).
ACCESSION   AJ224125
VERSION     AJ224125.1  GI:5420045
KEYWORDS    ERG gene; transcription factor.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1798)
  AUTHORS   Baltzinger,M., Mager Heckel,A.M. and Remy,P.
  TITLE     Xl erg: expression pattern and overexpression during development
            plead for a role in endothelial cell differentiation
  JOURNAL   Dev. Dyn. 216 (4-5), 420-433 (1999)
  MEDLINE   20099678
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Baltzinger,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1998) Baltzinger M., UPR9005, MMDCD, Centre
            National de la Recherche Scientifique, 15, RUE Rene Descartes,
            67084, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1798
```

```
                     /organism "Xenopus laevis"
                     /db_xref "taxon:8355"
                     /clone "lambda ZAP11 erq A"
                     /clone_lib "lambda ZAP11"
                     /dev_stage "stage 40 embryo"
     gene            321..1778
                     /gene "erq"
     CDS             321..1778
                     /gene "erq"
                     /function "transcription factor"
                     /codon_start=1
                     /protein_id "CAB46566.1"
                     /db_xref "GI:5420046"
                     /db_xref "SPTREMBL:Q9W700"
                     /translation "MQHMAFIPTYEDYSNEALSVVSEDQSLFECIYGTPHLIKTEMFA
                     SSSSDYGQTSKMSPRVPQQQWLSQPPSRVTIKMECNPSQVNGSRSSTPLMCSTGAGSKM
                     GGGSDNVGMNYGSYTEEKHTSPPNMTTNERRVTVPADPTLWSIDHVRQWLEWATKEYG
                     LPDVDVLLFQNIDGKELCKMTKEDFHRLTPSYNADILLSHLHYLRETPLPHLTSDDVD
                     KALQNSPRLMHARNTGGASFIFPNSSVYQDANQRITPSRQDLSYEPSRRSAWTNHPAPP
                     SKASQPSTLVPKTEIQPRPQLDPYQITGPTSSRLANPGSGQTQIWQFLLTLLTSDSSNSN
                     CITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRY
                     AYKFDFHGIAQALQPHPPESSLMYKYPSELPYMSSYHAHPQKMNFVAPHPPALPVTSSS
                     FFAAPNAYWNSPTGSTYPNTRLPASHMSSHLGTYY"
BASE COUNT       561 a    429 c    458 g    450 t
ORIGIN


alignment scores:
            Quality:  297.00              Length:     861
              Ratio:  1.727                 Gaps:      13
 Percent Similarity:  47.645   Percent Identity: 26.593

alignment block:
 US-09-126-945B-2 x XLAJ4125  ..

 Align seq 1/1  to: XLAJ4125  from: 1  to: 1798

      67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 88
         |||   :::   |||:::   :::   ::::::|||      :::
     468 TATGGGCAAACCTCGAAAATCAGTCCACGTGTGCCTCAACAAGACTGGCT 517

      83 qGluGluProPro..................................... 87
         ::::::||||||
     518 CTCACAGCCTCCATCCAGGGTGACCATCAAGATGGAATGCAACCCTAGTC 567

      88 ................GluGluProGluGlnCysProValIleAspSer 98
                        ::::::|||::::::|||   :::      :::
     568 AAGTCAATGGATCAAGGAGCTCCCCGGATGACTGCAGTATAGGAAAAGGA 617

      99 GlnAlaProAlaGlySerLeuAspLeuValPro.........GlyGlyLe 112
         :::      :::|||:::   |||   |||                |||:::
     618 AGCAAAATGGGTGGAGGGTCGGATAATGTTGGAATGAATTATGGCAGCTA 667

     112 qThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyG 129
             |||::|||          :::      :::|||:::::::::::|
     668 CATAGAAGAAAAACACATT......TCTCCCCCAAACATGACAACCAATG 711

     129 luValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAla 145
         ||                    :::   :::   :::   |||
     712 AG.......................AGGAGAGTGATTGTACCTGCT 744

     146 AspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpTh 162
         ||||||      ||||||   ::::::|||::::::|||||   ||:::
     745 GACCCTACTTTGTGGAGCACAGATCATGTCAGACAGTGGCTAGAGTGGGC 794

     162 rGlnHisGlnTyrArgLeuProPro.MetGly...LysAlaPheGlnGlnA 179
         :      :::|||   ||||||   :::           ||||||:::
     795 AATAAAAGAATACGGCCTTCCAGATGTCGATGTGCTCCTGTTCCAAAACA 844

     178 snAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgLnArg 194
         ::   ||||||||||||||||   |||::::::|||:::|||::::::
     845 TCGACGGCAAGGAATTATGCAAGATGACCAAAGAAGATTTCCATCGACTC 894

     195 SerPro...LeuGlyGlyAspValLeuHisAlaHisLeu..... 206
         :::|||      ::::::|||:::|||   :::||||||
     895 ACGGCAAGCTACAATGCTGATACCTTCTCTCGCATCTTCACTACCTCAG 944

     207 ..............................AspIleTrpLys..... 210
                                       ||::::   |||
     945 AGAAACTCCTCTTCCACATTTGACTTCAGATGATGTTGATAAGGCCTTAC 994

     211 .....SerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIleHis 225
              |||         |||      |||:::::   ||||||
     995 AAAACTCTCCACACGTTAATGCATGCTAGAAATACAGGAGGTGCCTCTTTT 1044

     226 TyrCysAlaSerThrSer............................... 231
                 :::::|||
    1045 ATTTTCCCAAACTCATCCGTTTATACAAGACGCAAACCAAAGAATCCCCAG 1094

     232 ..................................GluGluSerTrpThrAspS 248
                                          :::::::::||||||:::
    1095 CAGGCAAGATTTATCATATGAGCCCTCAAGGAGATCTGCTTGGACAAATC 1144

     248 erGlnValAspSerSerCysSerGlyGlnIro................... 248
           :::      |||   :::::|||||||
    1145 ATCCTGCAGCAGCATCAAAAGCATCTCAACCATCTACCACAGTTCCCAAA 1194

     248 ..................................................... 248

    1195 ACAGAAGACCCCAGGCCACAGCTAGATCCTTATCAGATTCTTGGGCCAAC 1244

     249 ..................................IleHisLeuTrpGlnP 264
                                          |||:::||||||||||
    1245 AAGCAGCCGACTTGCAAATCCAGGAAGTGGGCAGATTCAGCTCTGGCAAT 1294

     264 hTrSerLysGlnSerGlnLeuLysProHisSerTyrGlyArgPheIleAr 270
         |||||   |||   ||||||:::      |||   :::::   |||
    1295 TCTTACTGGAA...CTGCTTTCGGATAGTTCCAACTCCAACTGCATTACT 1341

     271 TrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVa 287
         |||   :::      ::::||   |||||||:::   |||      ::::||
    1342 TGC...GAAGGAACCAATGGAGAATTTAAGATGACCGATCCCGATGAAGT 1378

     287 lAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspL 304
         |||||||   ||||||   ||||||:::::::|||   ||||||||||||
    1379 GGCAAGACGTTGGGGGCAGAGGAAAAGCAAACCCAACATGAACTATGACA 1428

     304 ysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleIleArgLys 320
         ||||||||||||:::::|||   ||||||   |||::::::::   |||
    1429 AGCTCAGCCGTGCACTTCGTTACTACTACGATAAAAATATTATGACTAAA 1478

     321 ProAspIleSerGlnArgLeuValTyrGlnPhe 341
                 :::::::||   :::|||:::|||
    1479 GTCCAT...GGCAAGCGCTATGCTTACAAATTT 1508

seq_name: qb_pr5:AF115403

seq_documentation_block:
LOCUS       AF115403     2287 bp    mRNA            PRI       06-OCT-1999
DEFINITION  Homo sapiens Ets transcription factor ESE-2b mRNA, complete cds.
ACCESSION   AF115403
VERSION     AF115403.1  GI:4559274
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2287)
  AUTHORS   Seth,P., Kas,K., Jabs,A., Qi,X., Grill,F., Thangaraju,K.,
            Akbarali,Y., Finger,E., Boltax,J., Endress,G., Munger,K., Kunsch,C.
            and Libermann,T.A.
  TITLE     Characterization of ESE-2, a novel ESE-1 related Ets transcription
            factor that is restricted to glandular epithelium and
            differentiated keratinocytes
  JOURNAL   J. Biol. Chem. 274 (41), 29439-29452 (1999)
  MEDLINE   99436157
```

```
REFERENCE   2  (bases 1 to 2287)
  AUTHORS   Oettgen,P., Thamrongsak,U., Kas,K., Dube,A., Akbarali,Y.,
            Boltax,J., Kunsch,C., Endress,G. and Libermann,T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..2287
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p14"
                     /note="found in prostate and salivary glands"
     CDS             97..864
                     /codon_start=1
                     /product="Ets transcription factor ESE-2b"
                     /protein_id="AAD22961.1"
                     /db_xref="GI:4559275"
                     /translation="MESEVTHSTFLFNASFCEFLMSWTEKNGSYPTAFEHLTAVA
                     YWTSVHPEYWTKRHVWEWLQFCMDQYKLDINCISFCNENISGLQLCSM
                     IQEEFVEAAGLGFKLYFILQNIRTGQVSFENDAFFSKATIKDYAESNCLKISCIFSQLG
                     QSSHLWEFVRDLLLLSPEENCGILEWEDREQGIFRVVKSEALAEMWG
                     LSRALRYVYYKTGIIERVDRRLMYEPCENAHCWLEEL"
     repeat_region   1406..1505
                     /rpt_family="Alu"
BASE COUNT      666 a    488 c    478 g    655 t
ORIGIN


alignment_scores:
            Quality:  297.00              Length:     262
              Ratio:  1.993                 Gaps:       6
 Percent Similarity:  56.870   Percent Identity: 29.771

alignment_block:
 US-09-126-945B-2 x AF115403  ..

 Align seq 1/1  to: AF115403  from: 1  to: 2287

     102 AlaGlySerLeuAspLeuValProGlyGlyLeuThrLeu..........  141
         ||:::|||||||||||||||:  :::::|||||||:::
      58 GCCTCCTCTTTGGACCTAGCCACCGCTGCCCTCACGGTAATGTTGGCACTC  107

     115 .GluGluHisSerLeuGlnGlnValGlnSerMetValValGlyGlyVal  161
          :||||          ::::::                    ::::
     108 GGTGACACACAGCACCTTCCTGCCTAATGCATCCTTCTGCGATCCCCTGA  157

     131 euLys.....................................Asp  133
         :::::                                    :::
     158 TGTCGTGGACTGATCTGTTCAGCAATGAAGAGTACTACCCTGCCTTTCAC  207

     134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIl  183
         ::::|||||||              ::::::  |||     ||
     208 CATCAGACAGCCTGTGACTCATACTGGACATCAGTCCACCCTGAAIACTG  257

     150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrA  199
         ::::      ::::||   :::||||||   :::        ||||||
     258 GACTAAGCCCCATGTCTGGGAGTGGCTCCAGTTCTGCTGCCACCAGTACA  307

     167 rgLeuProProMetGlyLysAlaPhe......GluGluLeuAlaGlyLys  181
         ::|||           :::|||         ::::::|||
     308 AGTTGGACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGCCCTC  357

     182 GluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuCl  199
         :::|||||:::|||:::::|||:::|||   :::   :::   |||
     358 CAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAGCTGGCCTCTG  407

     198 yGlyAspValLeuHisAlaHisLeu.....................  206
          ||||::   |||:::     |||
     408 CGGCCACTACCTGTACTTCATCCTCCAGAACATCCCCACACAACGCTTACT  457

     207 ..........AspIleTrpLysSerAlaAlaTrpMetLysGlnArgThr  219
                  |||      :::|||   |||   :::|||:::   :::

     458 CCTTTTTTAATGACGCTGAAGAAAGCAAGGCCACCAACAAAGACTATGCT  507

     220 SerProGlyAlaIleHisTyrCysAlaSerThrSerGlnGluSerIrpTh  236
         :::   :::             |||   :::|||||      :::
     508 GATTCCAAC............TGCTTGAAAACAAGTGGCATCAAAAGTCA  545

     236 rAspSerGlnValAspSerSerCysSerGlyGlnProIleHisLeuTrpG  253
          |||   :::     |||     |||   |||      ||||||||:
     546 AGACTGTCACACTCTATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGG  595

     253 lnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle  269
         :|||:::::::::::|||||||||:::|||:::::        ::::::
     596 AATTTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTG  645

     270 AraTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl  286
         :::|||   :::::::|||:::||||||||||:::::      |||
     646 GAATGGGAAGATAGGAAACAAGGAATTTTCGGGTGGTTAAATCGGAAGC  695

     286 nValAlaArgLeuIrpGlyIleArgLysAsnArgProAlaMetAsnTyrA  303
         :::|||:::::::|||||   |||||||:::::      |||:::|||:
     696 CCTGCCAAACATCTCCCGACAAACGAACAAAAATCACACAAATGACTATC  745

     303 spLysLeuSerArgSerIleArgGlnTyrIyrLysLysGlyIleIleArg  319
         ::|||||||||||||:::::|||   |||||||||   ||||||:::::
     746 AAAAGTTGAGCAGAGCCCTGACATACTACTATAAAACAGGAATTTTGGAG  795

     320 LysProAspIleSerGlnArgLeuValIyrGlnPhe  331
         :::   |||      :::|||||:::|||:::|||
     796 CGGGTTGAC......CGAAGCTTAATGTACAAATTT  825

seq_name: qb_pr5:AF115402

seq_documentation_block:
LOCUS       AF115402      2452 bp    mRNA            PRI       06-OCT-1999
DEFINITION  Homo sapiens Ets transcription factor ESE-2a mRNA, complete cds.
ACCESSION   AF115402
VERSION     AF115402.1  GI:4559272
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2452)
  AUTHORS   Oettgen,P., Kas,K., Dube,A., Gu,X., Grall,F., Thamrongsak,U.,
            Akbarali,Y., Finger,E., Boltax,J., Endress,G., Munster,K., Kunsch,C.
            and Libermann,T.A.
  TITLE     Characterization of ESE-2, a novel ESE I-related Ets transcription
            factor that is restricted to glandular epithelium and
            differentiated keratinocytes
  JOURNAL   J. Biol. Chem. 274 (41), 29439-29452 (1999)
  MEDLINE   99436157
REFERENCE   2  (bases 1 to 2452)
  AUTHORS   Oettgen,P., Thamrongsak,U., Kas,K., Dube,A., Akbarali,Y.,
            Boltax,J., Kunsch,C., Endress,G. and Libermann,T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..2452
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p14"
                     /note="found in prostate and salivary glands"
     CDS             242..1029
                     /codon_start=1
                     /product="Ets transcription factor ESE-2a"
                     /protein_id="AAD22960.1"
                     /db_xref="GI:4559274"
                     /translation="MESEIPHSHRVMLESVTHSTFLFNASFCDFLMSWTEKESNEEYYP
                     AFEHQIACDSYWTSVHPEYWFKRHVWEWLQFCMDQYKLDINCISFCNENISGLQLCSM
                     IQEEFVEAAGLGFKLYFILQNIRTGQVSFENDAFFSKATIKDYAESNCLKISCIFSQLG
                     DCHSHSRTSLQSSHLWEFVRDLLLLSPEENCGILEWEDREQGIFRVVKSEALAKMWGQK
```

```
                    FFNEGMCYLEEGSALEYYYEICLLSVGEELMYEGCRAs WGLEL"
     repeat_region   1571. .1670
                     /rpt_family "Alu"
BASE COUNT       695 a    581 c    933 g    693 t
ORIGIN


alignment scores:
            Quality:  296.00              Length:    211
              Ratio:  2.295                 Gaps:      4
 Percent Similarity:  61.147   Percent Identity: 42.701

alignment block:
 US-09-126-945B-2 x AFI15402   ..

 Align seq 1/1  to: AFI15402  from: 1  to: 2452

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 184
          :::|||||||||||                :::::   |||      ||||::
      476 CAGACAGCCTGTGACTCATACTGGACATCAGTCCACCCTGAATACTGGAC 425

      185 rProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 234
          :      :::|||   :::|||||||   :::          ||||||||:|
      426 FAAGCGCCATGTGTGGGAGTGGCTCCAGTTCTGCTGCGACCAGTACAAGT 377

      235 euProProMetGlyLysAlaPhe.....GlnGluLeuAlaGly!, 276
          ||                :::|||         ::::::::|||   :::
      376 TGGACACCAATTGCATCTCCTTCTGCAACTTCAACATCAGTGGCCTGCAG 327

      277 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGl 326
          |||||||::|||::::::::|||:|||   :::   :::   |||   ||
      326 CTGTGCAGCATGACACAGGAGGAGTTCGTCGACGCAGCTGGGCCTCTGCGG 277

      327 yAspValLeuHisAlaHisLeu........................... 346
          |:::   ||||::       |||
      276 CGAGTACCTGTACTTCATCCTCCAGAACATCCGCACACAAGGTTACTCCT 225

      347 .......AspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSer 380
                |||       :::|||   |||   :::|||:::   ::::::
      226 TTTTTAATGACGCTGAAGAAAGCAAGCCCACCATCAAAGACTATGCTGAT 175

      381 ProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAs 430
          :::              |||   :::|||||       :::       ||
      176 TCCAAC............TGCTTGAAAACAAGTGGCATCAAAGTCAAGA 124

      431 pSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnP 480
          |   :::     |||      |||   |||      |||||||||||:|
      124 CTGTCACAGTCATAGTAGAACAAGCCTCCAAAGTTCTCATCTATGGGAAT 75

      481 heLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleAr 530
          ||:::::::::|||||||||:::|||:::::          :::::::::
      74  TTGTACAGACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATTCTGGAA 25

      531 TrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVa 580
          |||   ::::::||:::|||||||||||:::::      |||      ::
      24  TGGGAAGATAGGAACAAGGAATTTTTCGGGTGGTTAAATCGGAAGCCCT 75

      581 lAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspL 630
          :|||:::::|||||   |||||||:::::      |||:::|||:::|
      76  GGCAAAGATGTGGGGACAAAGGAAGAAAAATGACAGAATGACGTATGAAA 125

      631 ysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLys 680
          ||||||||||:::::::||   ||||||||   ||||||:::::::::::
      126 AGTTGAGCAGCCCTGAGATACTACTATAAAACAGGAATTTTGGAGCGG 175

      681 ProAspIleSerGluArgLeuValTyrGlnPhe 711
          |||       :::|||||:::::::::|||
      174 GTTGAC.....CGAAGGTTAATGTACAAATTT 146

seq name: qb_pat2:E31253

seq documentation block:
LOCUS       E31253       1447 bp    DNA             PAT       07 FEB 2001
DEFINITION  Protein having cell calcifying inhibitory activity and gene
            encoding the same.
ACCESSION   E31253
VERSION     E31253.1  GI:13625685
KEYWORDS    JP 1999075871 A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1447)
  AUTHORS   Hiroyasu,I.Y.H.B., Pashifiki,J.K. and Helena,E.
  TITLE     Protein having cell calcifying inhibitory activity and gene
            encoding the same
  JOURNAL   Patent: JP 1999075871 A 1 23-MAR-1999;
            CHUGAI PHARMACEUT CO LTD, BE TRUSTEES OF THE UNIVERSITY OF
            PENNSYLVANIA
COMMENT     OS   Unidentified
            PN   JP 1999075871 A/1
            PD   23-MAR-1999
            PF   29-MAY 1998 JP 1998166076
            PR   18-JUN 1997 US    08/878177,20-JUN 1997 US    60/050297 PI
            HIROYASU IWAMOTO,YOSHINOBU HIGUCHI,MARUJIO PASHIFIKI, PI  JOEL
            ROZENBUROOM,
            PI   HELENA E
            PC   C12N5/-, C A61K38/-, C07K14/-, C07K16/-, A61P35/00,
            PC   C12P21/02,
            FT   ,
            CC   Strandedness: Single:
            CC   Topology: Linear:
            FH   Key              Location/Qualifiers
            FT   source           1. .1447
            FT                    /organism "Unidentified".
FEATURES             Location/Qualifiers
     source          1. .1447
                     /organism "unidentified"
                     /db_xref "taxon:32644"
BASE COUNT       440 a    474 c    317 g    316 t
ORIGIN


alignment scores:
            Quality:  294.50              Length:    346
              Ratio:  1.753                 Gaps:     14
 Percent Similarity:  48.555   Percent Identity: 27.457

alignment block:
 US-09-126-945B-2 x E31253   ..

 Align seq 1/1  to: E31253  from: 1  to: 1447

       67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 83
          |||   :::   ||||::   :::   :::::|||      :::
      183 TATGGGCAAACATCAAAGATGACCCCGCGGTCCCCAGCAGGACTGGTT 232

       84 qGluGluProPro..................................... 87
          ::::::||||||
      233 ATCACAGCCCCCGGCCAGAGTTACCATTAAGATGGAGTCTAACCCAAACC 282

       88 ..................GluGluProGluGlnCysProValIleAspSer 98
                          ::::::|||::::::|||   |||      :::
      283 AGGTTAATGGGTCAAGGAATTCACCTGATGACTGCAGCGTGGCAAAAGGA 332

       99 GluAlaProAlaGlySerLeuAspLeuValPro.........GlyGlyLe 112
             :::::|||   |||   |||                 ||||::
      333 GGGAAAATGGTTAGCAGTTCAGACAATGTTGGGATGAACTATGGAAGCTA 382

      112 uThrLeuGluGluHisSerLeuLeuGlnValGlnSerMetValValGlyG 129
             ||||:|||            :::|||:::::::::
      383 CATGGAACAGAAGCATATT.....CCGCCTCCAAATATGACAACCAATG 426

      129 luValIleLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAla 145
          ||                        :::   :::   :::   |||
      427 AA........................CGAAGAGTTATTGTGCCAGCA 449
```

```
     146 AspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpIh 162
         ||||||      ||||||   :::::||||::::::|||||   ||||:
     450 GATCCTACGTTATGGAGCACAGACCATGTACGGCAGTGGCTGGAGTGGGC 499

     162 rGluHisGlnTyrArqLeuProProMetGly...LysAlaPheGluGluL 178
         :      :::|||   ||||||   :::           |||||||:::
     500 AGTGAAGGAGTATGGTCTTCCAGACGTGGACATCTTGTTGTTCCACAACA 549

     178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArqGlnArq 194
         ::   |||||||||||||||   |||:::::::::::::|||::::::
     550 TTGATGGCAAAGAGTTGTGTAAAATGACCAAAGATGACTTCCAGAGACTC 599

     195 SerPro...LeuGlyGlyAspValIeuHisAlaHisLysAspIleTrpTy 213
         :::|||      :::::|||:::|||   :::|||||
     600 ACGCCGAGCTATAACGCAGATATCCTCCTGTCACACCTACAC..... 641

     213 sSerAlaAlaTrpMetLysGluArqThrSerProGlyAlaIleHis3,rC 227
                :::::::::::|||||         ||||||
     642           TACCTCAGAGAGACA         GGAGCCACTTTTATTT 672

     227 ysAlaSerThrSer                                   231
         :::|||||||
     673 TTCCAAATACATCAGTTTACCCAGAAGCAACGCAAACAATAACAACAAGG 722

     232 ..........................GluGluSerTrp........... 245
                                   ::::::::|||
     723 CCACATTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGACGAGTCACAG 772

     235 ..................................................... 235

     773 CCATCCCACTCAGTCAAAAGCTACCCAACCATCATCTTCAACAGTGGCCA 822

     235 ..................................................... 235

     823 AAACAGAAGACCAGCGTCCTCAGTTAGATCCTTATCAGATTCTTGGACCG 872

     236 ThrAspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTr 252
         |||::::|:::::::   :::     ||||||||   |||:::||||
     873 ACCAGCAGCCCTCTTGCAAATCCAGGGAGTGGGCAG...ATACAGCTATG 919

     252 pGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArqPheI 269
         |||||||:||   |||   ||||||:::      |||   ::::::   |
     920 CCACTTCCTACTGGAG...CTTCTGTCGGACAGCTCCAACTCCAACTGCA 968

     269 leArqTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAla 285
         ||   ||:   :::      :::|||   ||||||:::   |||
     967 TCACCTGG...GAGGGCACAAATGGGCAGTTCAAGATGACAGACCCTGAT 1013

     286 GluValA.aArqLeuTrpGlyIleArqLysAsnArqProAlaMetAsnTy 302
         :::||||||||   ||||||   |||||:::::::|||   ||||||||
    1014 GAAGTGGCTCGGCGTTGGGCAGAGAGGAAAAGCAAACCTAACATGAACTA 1063

     302 rAspLysLeuSerArqSerIleArqGlnTyrTyrLysLysGlyIleIleA 319
         ||||||||||||||||:::::|||   ||||||   |||:::|||:::
    1064 TGACAAACTCAGCCGTGCACTTCGCTACTACTATGACAAAAATATTATCA 1113

     319 rqLysProAspIleSerGlnArqLeuValTyrGlnPhe 331
           |||         :::::|||   :::|||:::|||
    1114 CTAAAGTTCAT...GGTAAACGCTATGCCTACAAATTT 1148

seq_name: qb_pr10:HSU13044

seq_documentation_block:
LOCUS       HSU13044     1970 bp    mRNA            PRI       03-MAY-1995
DEFINITION  Human nuclear respiratory factor-2 subunit alpha mRNA, complete
            cds.
ACCESSION   U13044
VERSION     U13044.1  GI:531892
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo

REFERENCE   1  (bases 1 to 1970)
  AUTHORS   Gugneja,S., Virbasius,J.V. and Scarpulla,R.C.
  TITLE     Four structurally distinct, non DNA binding subunits of human
            nuclear respiratory factor 2 share a conserved transcriptional
            activation domain
  JOURNAL   Mol. Cell. Biol. 15 (1), 102-111 (1995)
  MEDLINE   95097980
REFERENCE   2  (bases 1 to 1970)
  AUTHORS   Gugneja,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-AUG-1994) Sajiv Gugneja, Cell and Molecular Biology,
            Northwestern University, 303 E. Chicago Ave, Chicago, IL 60611, USA
FEATURES             Location/Qualifiers
     source          1. .1970
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HeLa"
     CDS             186. .1550
                     /citation=[1]
                     /codon_start=1
                     /product "nuclear respiratory factor 2 subunit alpha"
                     /protein_id "AAA65706.1"
                     /db_xref="GI:531893"
                     /translation="MTKPEAEFLIFIFIDGIFKAFCTFPSIVFQTYAPAFCVSQAIDI
                     NEPIGNEKKLLEFKLQCSLDAIHFICLQDIQLDPERSLFDQGVKIDGTVQLSVQVISYQ
                     GIEPKLNILEIVKPADIVFVVIPPAHHAESEAHLVEFAQVTIIIGIKPIIFLSDFIS
                     EQVTPWAAALEGYKKEQEPLGIPYDPIQWSTLQVIHWVVWVMKEFSMIDIDLITLNIS
                     GRELCSLNQEDFFQRVPRGEILWSHLELLKKYVLASQEQQMNEIVTIDQPVQIIPASV
                     QSATPITIKVINRCAKAAKVQKAPKISGLDRSSPGNRTGNNGQTQLWQFILELLSDKD
                     APICISWVGDFCFFKLNQFELVAQKWKQKKNKPIMNYEKLSRALRYYYIAGIMICKVQG
                     FPPVYKEFVCDLETLICYSAAFLNPIVTFCTPQFKLAKMQLHGIAQPVTVVALAIASLQT
                     EKDN"
BASE COUNT      618 a    382 c    457 g    513 t
ORIGIN


alignment_scores:
           Quality:  293.00            Length:    306
             Ratio:  1.694               Gaps:      9
 Percent Similarity:  56.536   Percent Identity: 28.105

alignment_block:
 us-09-126-945b-2 x HSU13044   ..

 Align seq 1/1  to: HSU13044  from: 1  to: 1970

      60 LeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLy 76
         |||:::   :::::::::::   |||   |||:::      ::::::
     507 TTAAACATCCTTGAAATTGTTAAACCTGCGGACACTGTTGAGGTTGTTAT 556

      76 sAlaProGlyAlaSerSerArqGluGluProProGluGluProGluGlnC 93
          |||   |||         |||:::      :::      |||:::
     557 TGATCCAGATGCCCACCATGCTGAATCAGAAGCACATCTTGTTGAAGAAG 606

      93 ysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValPro 109
         ::   ||||||                  :::||||||   :::
     607 CTCAAGTGATA.....................ACTCTTGATGGCACAAAA 645

     110 GlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetVa 126
               |||:::::::::::::::::   |||||||||         ::
     636 CACATCACAACCATTTCAGATGAAACTTCAGAACAAGTGACAAGATGGGC 685

     126 lValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnI 143
         ::::      :::|||:::      :::         :::   |||:::|
     686 TGCT......GCACTCGAACGCTATACGAAAGAACAAGAACGCCTTGGCA 729

     143 leThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeu 159
         ||      |||||:::::::||||||   ::::::|||      |||:::
     730 TACCCTATGATCCCATACAGTGGTCCACAGACCAAGTCCTGCATTGGGTG 779

     160 LeuTrpThrGluHisGlnTyrArqLeuProProMetGlyLysAlaPheGl 176
         :::|||:::      :::::   :::      :::      :::
     780 GTTTGGGTAATGAAGGAATTCAGCATGACCGATATAGACCTCACCACACT 829
```

```
     176 nGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGlnGlnPheArgG 193
         ::::::::::|||::|||||||||||:::::::::::|||:::|||   |
     830 CAACATTTCGGGGAGAGAATTATGTAGTCTCAACCAAGAAGATTTTTTC 879

     194 lnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrp 209
         |||||   |||       |||::::::|||   :::|||||||:::::
     880 AGCGGGTTCCT...CGGGGAGAAATTCTCTGGAGTCATCTGGAACTTCTC 926

     210 LysSerAlaAlaTrpMetLysGluArg..................... 218
         ::::::   :::       ::::::::
     927 CGAAAATATGTATTGGCAACTCAAGAACAACAGATCAATGAAATAGTTAC 976

     218 ..................................................... 218

     977 AATTGATCAACCTGTGCAAATTATTCCAGCATCAGTGCAAICTGCTACAC 1026

     219 ..ThrSerProGlyAlaIleHisTyrCysAlaSerThrSer......Glu 232
         |||:::      :::|||:::   |||||||:::::::      :::
    1027 CTACTACCATTAAAGTTATAAATAGATGTCCGAAAGCAGCCAAAGTACAA 1076

     233 GluSerTrpThrAspSerGluValAspSerSerCysSerGlyGln..... 247
         ::::::         |||       |||   |||       |||:::
    1077 AGAGCGCCGAGGATTTCAGGAGAAGATAGAAGCTCACCTGGGAACAGAAC 1126

     248 .............ProIleHisLeuTrpGlnPheLeuLysGlnLeuLeuL 269
                     |||:::|||||||||||||||   |||||||||
    1127 AGCAAACAATGGCCAAATCCAACTATGGCAGTTTTGCTAGAACTTCTTA 1176

     269 euLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLys 276
              :::                |||   |||::::::   |||
    1177 CTGATAAGGACGCTCGAGAC...TGCATTTCTTGGGTTGGTGATGAA... 1220

     277 GlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyLl 293
         |||   ||||||:::::::::          ||||||:::   ||||||
    1221 GGTGAATTTAAGCTAAATCAGCCTGAACTGGTTGCACAGAAATGGGGACA 1270

     293 eArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleA 310
         ||||||||||:::||:::||||||||||:::||||||||||||:::::|
    1271 GCGTAAAAATAAGCCTACGATGAACTATGACAAACTCAGTCGTGCATTAA 1320

     310 rgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArg 326
         ||   ||||||          :::|||   |||   :::   :::::|||
    1321 GATATTATTACGATGGGACATGATTTGTAAAGTTCAA...GGCAAGAGA 1367

     327 LeuValTyrGlnPheVal 332
         :::|||||||:::||||||
    1368 TTTGTGTACAAGTTTGTC 1385

seq name: qb pr5:AK026144

seq documentation block:
LOCUS       AK026144     1736 bp    mRNA            PRI       29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ22491 fis, clone HRC11042, highly similar to
            AF116509 Homo sapiens Ets transcription factor TEL-2b (TEL-2) mRNA.
ACCESSION   AK026144
VERSION     AK026144.1  GI:10438901
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens primary human renal epithelial cells cDNA to mRNA,
            clone lib:HRC clone:HRC11042.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1736)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Inagaki,H.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center,
            Shirokanedai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/qualifiers
     source          1..1736
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="primary human renal epithelial cells"
                     /clone="HRC11042"
                     /clone_lib="HRC"
                     /note="cloning vector pME18SFL3"
     misc_feature    1..1736
                     /note="highly similar to AF116509 Homo sapiens Ets
                     transcription factor TEL-2b (TEL-2) mRNA"
BASE COUNT       413 a    508 c    515 g    300 t
ORIGIN


alignment scores:
            Quality:  292.50            Length:    403
              Ratio:  1.556               Gaps:     14
 Percent Similarity:  46.650   Percent Identity:  24.566

alignment_block:
 US-09-126-945B-2 x AK026144   ..

 Align seq 1/1  to: AK026144  from: 1  to: 1736

       6 ProGlyLeuSerSerValSerProSerHisLeuLeuLeuProProAspTh 22
         ||||||   :::   :::|||   |||:::      :::|||||||
      61 CCCGGGCGCACACGCCTTAGCGCTTCTCGAAAGAGATTTCCTCCCACGCG 110

      22 rValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVal...GlyL 38
             |||                 ::::::   |||   |||   ||||
     111 ACCTTCC.................AGTTCTCGGGGCCAGGTTAGGGGTT 142

      38 euGluArgArgAspTrpSerProSerPro.....ProAlaThrProGlu 52
         ||   :::::|||   :::                |||::::::   :::
     143 TGGCGGAGGAGACTGCGGGGCGCGGGCCTAGGGCCCAGCAGCCACAGC 192

      53 GluGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGl 69
         ||||||
     193 CAGGGG..................................................... 198

      69 aAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluP 86
            |||          :::   |||   |||||||||          |
     199 .......AGCGCTCAAGACAGAAAGCCGGTGGCTTCCTCA.....CCTC 245

      86 roProGluGluProGluGlnCysProValIleAspSerGlnAlaProAla 102
         |||||          |||          :::|||:::      :::|||:::
     246 CACCTGTAATGCAGGAGGGAGAATTGGCTATTCTCCTATAAGCCCTGTG 295

     103 GlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLe 119
         ::::::::::          |||   |||       ::::::   :::
     296 GCAGCCATGCCT......CCCTAGGCACCCACGTGCAAGCCAGATGTGA 329

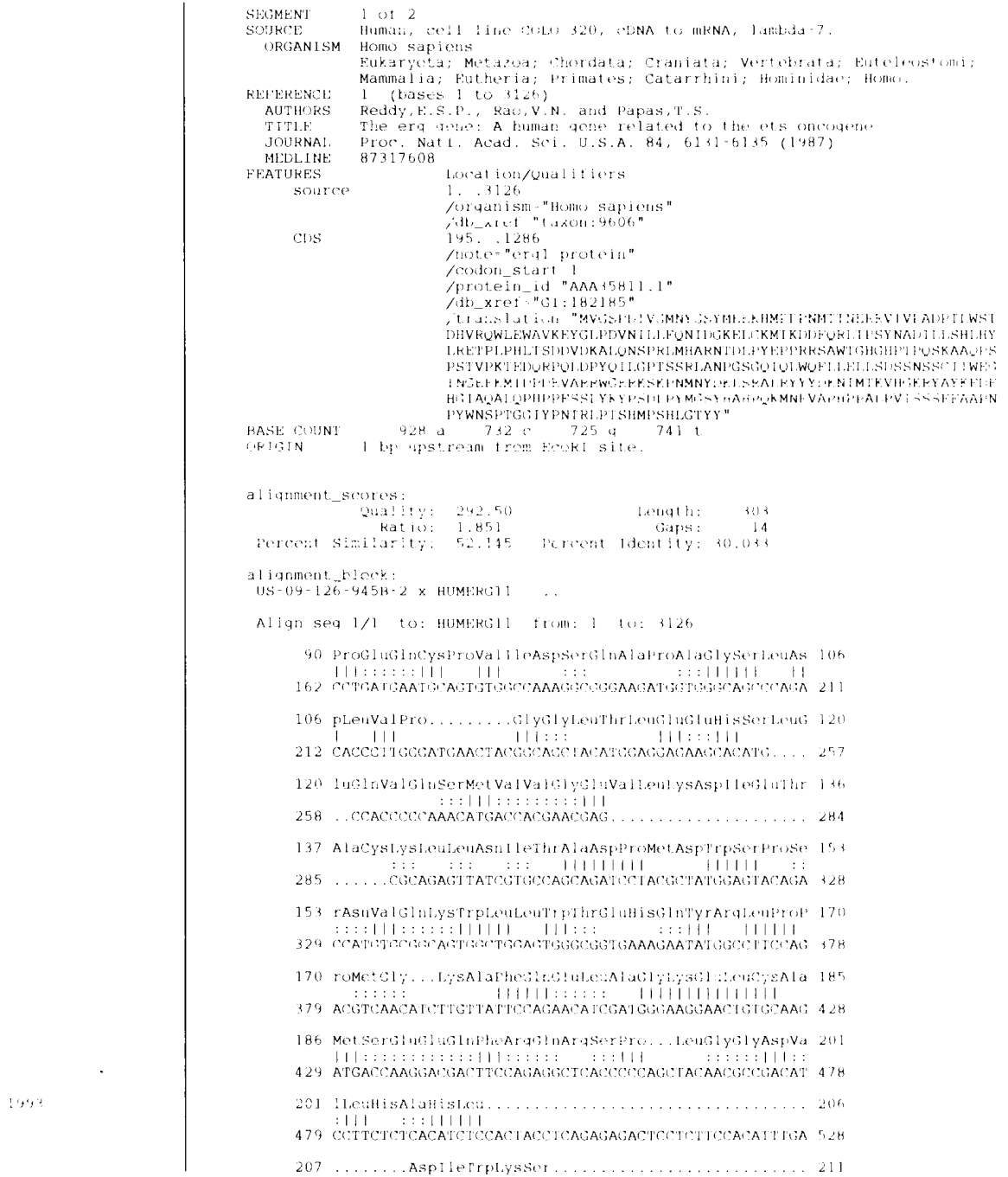
     119 uGlnGluValGlnSerMetValValGlyGluValLeuLysAspIleGluI 136
             |||::::::   :::       |||   :::
     330 AGCTCAAATTAACCTGCTGGGTGAAGGGAGGATC................ 363

     136 hrAlaCysLysLeuLeu...AsnIleThrAlaAspProMetAspIrpSer 151
         ||||||||||      ::::::      :::|||      ||||||
     364 .....TGCAAGCTGCCAGGAAGACTCCGCATCCAGCCCGCACTGTGGAGC 408
```



```
seq_name: qb_pr10:HUMERG11

seq_documentation_block:
LOCUS       HUMERG11     3126 bp    mRNA            PRI       27-APR-1993
DEFINITION  Human erg protein (ets-related gene) mRNA, complete cds.
ACCESSION   M21535 M17390
VERSION     M21535.1  GI:182182
KEYWORDS    erg protein.
```



```
SEGMENT     1 of 2
SOURCE      Human, cell line COLO 320, cDNA to mRNA, lambda-7.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3126)
  AUTHORS   Reddy,E.S.P., Rao,V.N. and Papas,T.S.
  TITLE     The erg gene: A human gene related to the ets oncogene
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 84, 6131-6135 (1987)
  MEDLINE   87317608
FEATURES              Location/Qualifiers
     source           1. .3126
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
     CDS              195. .1286
                      /note="erg1 protein"
                      /codon_start=1
                      /protein_id "AAA45811.1"
                      /db_xref="GI:182185"
                      /translation "MVGSPEEVGMNVGSYMLEKPMETSNMCTPGERSVTVPADPTLWSI
                      DDVRQWLEWAVKEYGLPDVNILLFQNIDGKELCKMTKDDFQRLTPSYNADILLSHLHY
                      LRETPLPHLTSDDVDKALQNSPRLMHARNTDLPYEPPRRSAWTGHGHPTPQSKAAQPS
                      PSTVPKTEDQRPQLDPYQILGPTSSRLANPGSGQIQLWQFLLELLSDSSNSSCITWEG
                      TNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDF
                      HGIAQALQPHPPESSLYKYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSFFAAPN
                      PYWNSPTGGIYPNTRLPTSHMPSHLGTYY"
BASE COUNT      928 a    732 c    725 g    741 t
ORIGIN      1 bp upstream from EcoRI site.
```



```
alignment_scores:
            Quality:  292.50              Length:    303
              Ratio:  1.851                 Gaps:     14
 Percent Similarity:  52.145   Percent Identity:  30.033

alignment_block:
 US-09-126-945B-2 x HUMERG11   ..

 Align seq 1/1  to: HUMERG11  from: 1  to: 3126
```

```
                 ||:::   |||:::
     529 CTTCAGATGATGTTGATAAAGCCTTACAAAACTCTCCACGGTTAATGCAT 578

     212 .....AlaAlaTrp.. 214
                    :::|||||
     579 GCTAGAAACACAGATTACCATATGAGCCCCCCAGGAGATCAGCCTGGAC 628

     215 .....MetLysGluArgThrSerPro 222
                       :::   |||||||
     629 CGGTCACGGCCACCCCACGCCCCAGTCGAAAGCTGCTCAACCAPCTGCTT 678

     223 TyAL:::HisTyrCysAlaSerLeuSerCysHis 244
         ::::::::            :::|||:::::::::
     679 CCACAGTG........CCCAAAACTGAAGACCAGCGTCCTCAGTTAGAT 719

     244 ..............SerTrpThrAspSerGluValAspSerSerCysSe 245
                      :::   |||:::|||:::::   :::      ||
     720 CCTTATCACATTTTTGCACCAACAACTACCCCCTTCCAAATCCACCCAC 769

     245 rGlyGluProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysP 262
         |||||||   |||:::|||||||||||||||   |||   |||||||::
     770 TGGCCAG...ATCCAGCTTTGGCAGTTCCTCCTGGAG...CTCCTGTCGG 813

     262 roHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
           |||   :::    |||   |||   :::      :::|||
     814 ACAGCTCCAACTCCAGCTGCATCACCTGG...GAAGGCACCAACGGGGAG 860

     279 PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLy 295
         |||||||:::   |||      :::||||||||||   ||||||   |||||
     861 TTCAAGATCGACGGATCCCGACCAGGTGGCCCGGCGCTGGGGAGAGCGGAA 910

     295 sAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnI 312
         |:::::|||   |||||||||||||||||||||||||:::::|||   |
     911 GAGCAAACCCAACATGAACTACGATAAGCTCAGCCGCGCCCTCCGTTACT 960

     312 yrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVal 328
         |||||   |||:::|||:::   |||         :::::|||   :::
     961 ACTATGACAAGAACATCATGACCAAGGTCCAT...GGGAAGCGCTACGCC 1007

     329 TyrGlnPhe 331
         |||:::|||
    1008 TACAACTTC 1016

Seq_name: qb ov:GGERG

seq_documentation_block:
LOCUS       GGERG          1516 bp    mRNA            VRT       27 APR 1995
DEFINITION  G.gallus ERG mRNA.
ACCESSION   X77159
VERSION     X77159.1  GI:790439
KEYWORDS    ERG gene.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1516)
  AUTHORS   Dhordain,P., Dewitte,F., Desbiens,X., Stehelin,D. and
            Duterque Coquillaud,M.
  TITLE     Mesodermal expression of the chicken erg gene associated with
            precartilaginous condensation and cartilage differentiation
  JOURNAL   Mech. Dev. 50 (1), 17-28 (1995)
  MEDLINE   95329425
REFERENCE   2  (bases 1 to 1516)
  AUTHORS   Duterque Coquillaud,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17 JAN 1994) M. Duterque Coquillaud, CNRS UA 1160,
            Oncologie Moleculaire, Institut Pasteur, 1 rue Calmette, 59019
            Lille, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1516
                     /organism "Gallus gallus"
                     /db_xref "taxon:9031"
                     /dev_stage "adult"
                     /tissue_type "spleen"
                     /clone_lib "lambda gt10"
     gene            68. .1499
                     /gene "ERG"
     CDS             68. .1499
                     /gene "ERG"
                     /codon_start=1
                     /protein_id "CAA54404.1"
                     /db_xref "GI:790440"
                     /db_xref "SWISS-PROT:Q90837"
                     /translation="MAEEFEIAESVVSEDQSLESCAYGGTPELAEETDMIASSSSEYGQI
                     SKMSPEVPQQSWLSQPPARVTLKMECNPSQVNGSRNSPDECSVAKGGKMVGSSDNVGM
                     NYGSYMEEKHIPPPNMTTNERKVIVPADPTLWSTDHVRQWLEWAVKEYGLPDVDILLF
                     QNIDGKELCKMTKDDFQRLTPSYNADILLSHLHYLRETPLPHLTSDDVDKALQNSPRL
                     MHARNTGGAEFIFPNTSVYPEATQRITTKPDLPYEQARRSAWTSHSHEPQSKATQPSS
                     STVIPKTEDQRPQLDPYQILGPTSSRLANPGSGQIQLWQFLLELLSDSSNSNCITWEGT
                     NGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFH
                     GIAQALQPHPPESSMYKYPSDLPYMSSYHAHPQKMNFVAPHPPALPVTSSSFFAAPNP
                     YWNSPTGGIYPNTRLPAAHMPSHLGTYY"
BASE COUNT      458 a     392 c     427 g     439 t
ORIGIN


alignment scores:
            Quality:  292.00              Length:     864
              Ratio:  1.708                 Gaps:      14
 Percent Similarity:  46.978   Percent Identity: 26.923

alignment block:
 US-09-126-945B-2 x GGERG   ..

 Align seq 1/1  to: GGERG  from: 1  to: 1516

      67 TyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerAr 83
         |||   :::   |||:::   :::   :::::::|||      :::
     188 TATGGGCAAACATCAAAGATGAGCCCGCGCGTTCCCCAGCAGGACTGGTT 232

      84 qGluGluProPro..... 87
         ::::::||||||
     233 ATCACAGCCCCCGGCCAGAGTTACCATTAAGATGGAGTGTAACCCAAACC 282

      88 .....GluGluProGluGlnCysProValIleAspSer 98
                ::::::|||:::::|||   |||      :::
     283 AGGTTAATGGGTCAAGGAATTCACCTGATGACTGCAGCGTGGCAAAAGGA 332

      99 GlnAlaProAlaGlySerLeuAspLeuValPro.....GlyGlyLe 112
               ::::::|||   |||   |||          ||:::
     333 GGGAAAAPGGTTAGCAGTTCAGACAATGTTGGGATGAACTATGGAAGCTA 382

     112 uThrLeuGluGluHisSerLeuGlnGlnValGluSerMetValValGlyG 129
           ||:::|||                 :::|||:::::::::::
     383 CATGGAAGAGAAGCATATT.....CCGCCTCCAAATATGACAACCAATG 426

     129 luValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAla 145
         ||                        :::   :::   :::   |||
     427 AA.....CGAAGAGTTATTGTGCCAGCA 449

     146 AspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpTh 162
         ||||||      ||||||   ::::::|||::::::{||||||   |||::
     450 GATCCTACGTTATGGAGCACAGACAATGTACGGCAGTGGCTGGAGTGGGC 499

     162 rGluHisGlnTyrArgLeuProProMetGly...LysAlaPheGlnGluL 178
         :      ::|||   ||||||   :::           ||||||||::
     500 AGTGAAGGACGTATGGTCTTCCAGACGTGGACATCTTGTTGTTGCAGAACA 549

     178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArg 194
         ::   |||||||||||||||   |||:::::::::::::|||::::::
     550 TTGATGGGAAAGAGTTGTGTAAAATGACCAAAGATGACTTCCAGAGACTC 599

     195 SerPro...LeuGlyGlyAspValLeuHisAlaHisLeu..... 206
         :::|||      ::::::|||:::|||(   :::||||||
     600 ACGCCGAGCTATAACGCAGATATCCTCCTGTCACACCTACACTACCTCAG 649
```

```
      207 ..............................AspIleIrpLys..... 219
                                       ||:::   |||
      650 AGAGACTCCTCTTCCACATTTGACTTCAGATGATGTTGATAAGGCCTAC 699

      220       SerAlaAlaTrpMetLysGluArgThrSerLeuArgArg..."
                ||| |||  |||  ||||||||  |||||||
      700 AAAACTCTCCACGGTTAATGCATGCTAGAAACACAGGAGGAGCCACTTTT 749

      226 TyrCysAlaSerThrSer.............................. 241
                ||||||||
      750 ATTTTTCCAAATACATCAGTTTACCCAGAAGCAACGCAAAGAATAACAAC 799

      242 ..............................GluGluSerIrp....... 255
                                       ||||||||||||
      800 AAGGCCAGATTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGACGAGTC 849

      256 ................................................. 285

      850 ACAGCCATCCCACTCAGTCAAAAGCTACCCAACCATCATCTTCAACAGTG 899

      286 ................................................. 315

      900 CCCAAAACAGAAGACCAGCGTCCTCAGTTAGATCCTTATCAGATTCTGG 949

      316 ....ThrAspSerGluValAspSerSerCysSerGlyGlnProIleHis. 351
              |||:::|||:::::  :::     |||||||||  |||:::|
      950 ACCGACCAGCAGCCGTCTTGCAAATCCAGGGAGTGGGCAG...ATACAGC 996

      352 euIrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArg 397
          ||||||||||||||  |||  ||||||||:     |||  :::::
      997 TATGGCAGTCCTACTGGAG...CTTCTGTCGGACAGCTCCAACTCCAAC 1043

      398 PheIleArgIrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSe 284
          |||  |||  :::     ::|||  ||||||::   |||
     1044 TGCATCACCTGC...GAGGCCACAAATGGGGAGTTCAAGATGACAGACGC 1090

      284 rAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetA 301
          :::|||||||||  ||||||  |||||||::::::||  ||||
     1091 TGATGAAGTGGCTCGGCCTTGGGAGAGAGGAAAAGCAAACCTAACAATGA 1140

      301 snTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIle 317
          ||||||||||||||||||:::::||  ||||||  |||:::|||
     1141 ACTATGACAAACTCAGCCGTGCACTTCGCTACTACTATGACAAAAATATT 1190

      318 IleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          :::  |||          :::::|||  :::|||:::|||
     1191 ATGACTAAAGTTCAT...GGTAAACGCTATGCCTACAAATTT 1229

seq_name: qb_pat2:E31254

seq_documentation_block:
LOCUS       E31254       1528 bp    DNA             PAT       07-FEB-2001
DEFINITION  Protein having cell calcifying inhibitory activity and gene
            encoding the same.
ACCESSION   E31254
VERSION     E31254.1  GI:13025686
KEYWORDS    'P 1999075871 A/2
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1528)
  AUTHORS   Hiroyasu,I.Y.H.H., Pashifiki,J.R. and Helena,E.
  TITLE     Protein having cell calcifying inhibitory activity and gene
            encoding the same
  JOURNAL   Patent: JP 1999075871-A 2 23-MAR-1999;
            CHUGAI PHARMACEUT CO LTD; HE TRUSTEES OF THE UNIVERSITY OF
            PENNSYLVANIA
COMMENT     OS   Unidentified
            PN   JP 1999075871 A/2
            PD   23-MAR-1999
            PF   29-MAY-1998 JP 1998166076
            PR   18-JUN-1997 US    08/878177,20 JUN 1997 US    08/050297 PI
            HIROYASU IWAMOTO,YOSHINOBU HIGUCHI,MARIJIO PASHIFIKI, PI    JOEL
            ROZENBURGOM,
            PI   HELENA E
            PI   [illegible]
            PC   C12P21/02,
            PI   [illegible]
            CC   Strandedness: Single
            CC   Topology: Linear
            FH   Key             Location/Qualifiers
            FT   source          1. .1528
            FT                   /organism "Unidentified".
FEATURES             Location/Qualifiers
     source          1. .1528
                     /organism "unidentified"
                     /db_xref "taxon:32644"
BASE COUNT      464 a     395 c     330 g     339 t
ORIGIN


alignment_scores:
            Quality:  292.00              Length:     364
              Ratio:  1.708                 Gaps:      14
 Percent Similarity:  46.978   Percent Identity: 26.923

alignment_block:
 US-09-126-945B-2 x E31254   ..

 Align seq 1/1  to: E31254  from: 1  to: 1528

       67 TyrProGluAspSerSerIrpAlaAlaLysAlaProGlyAlaSerSerAr 84
          |||  :::  ||:::  :::  ::::::|||  :::
      183 TATGGGCAAACATCAAAGATGAGCCCGCGCTTCCCCAGCAGGACTGGTT 232

       83 qGluGluProPro..................................... 87
          ::::::|||||
      233 ATCACAGCCCCCGGCCAGAGTTACCATTAAGATGGAGTGTAACCCAAACC 282

       88 ................GluGluProGluGlnCysProValIleAspSer 98
                          ::::::|||:::::|||  |||     :::
      283 AGGTTAATGGGTCAAGGAATTCACCTGATGACTCCAGCGTGGCAAAAGGA 332

       99 GlnAlaProAlaGlySerLeuAspLeuValPro.........GlyGlyLe 112
             ::::|||  |||  |||           ||:::
      333 GGGAAAATGGTTAGCAGTTCAGACAATGTTGGGATGAACTATGGAAGCTA 382

      113 uThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyG 129
             |||::|||           :::|||:::::::::|
      383 CATGGAAGAGAAGCATATT......CCGCCTCCAAATATGACAACCAATG 426

      129 luValIleLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAla 145
          ||                      :::  :::  :::  |||
      427 AA.....................CGAAGAGTTATTGTGCCAGCA 449

      146 AspProMetAspTrpSerProSerAsnValGluLysIrpLeuLeuIrpTh 162
          ||||||     ||||||  ::::::|||:::::|||||  ||:::
      450 GATCCTACGTTATGGAGCACAGACCATGTACGGCAGTGGCTGGAGTGGGC 499

      163 rGluHisGluTyrArgLeuProProMetGly  lysAlaPheGlnGluL 178
          :     :::|||  ||||||  :::          ||||||::::
      500 AGTGAAGGAGTATGGTCTCCAGACGTGGACATCTTGTTGTCCAGAACA 549

      178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArg 194
          ::  ||||||||||||||  |||:::::::::::|||:::::
      550 TTGATGGCAAAGAGTTGTGTAAAATGAGCAAAGATGACTTCCAGAGACTC 599

      195 SerPro...LeuGlyGlyAspValLeuHisAlaHisLeu.......... 206
          :::|||     ::::::|||:::|||   :::|||||
      600 AGGCCGAGCTATAACGGCAGATATCCTCCTGTCACACCTACACTACCTCAG 649

      207 ..............................AspIleIrpLys..... 219
                                       ||:::   |||
      650 AGAGACTCCTCTTCCACATTTGACTTCAGATGATGTTGATAAGGCCTAC 699
```

```
     211 ....SerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIleHis 225
             |||          |||      |||:::::   ||||||
     700 AAAACTCTCCACGGTTAATGCATGCTAGAAACACAGGAGGAGCCACTTTT 749

     226 TyrCysAlaSerThrSer..............................  231
             :::|||||||
     750 ATTTTTCCAAATACATCAGTTTACCCAGAAGCAACGCAAAGAATAACAAC 799

     232 ..............................GluGluSerTrp.......  235
                                        ::::::::::|||
     800 AAGGCCAGATTTACCTTATGAGCAAGCGAGGAGATCAGCGTGGACGAGTC 849

     235 ................................................  235

     850 ACAGCCATCCCACTCAGTCAAAAGCTACCCAACCATCATCTTCAACAGTG 899

     235 ................................................  235

     900 CCCAAAACAGAAGACCAGCGTCCTCAGTTAGATCCTTATCAGATTCTTGG 949

     236 ....ThrAspSerGluValAspSerSerCysSerGlyGlnProIleHisL 251
             |||:::|||::::::   :::      |||||||||   |||:::|
     950 ACCGACCAGCAGCCGTCTTGCAAATCCAGGGAGTGGGCAG...ATACAGC 996

     251 euTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArg 267
         ||||||||||||||   |||   |||||||:::      |||   ::::::
     997 TATGGCAGTTCCTACTGGAG...CTTCTGTCGGACAGCTCCAACTCCAAC 1043

     268 PheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSe 284
             |||   |||   :::      :::|||   ||||||:::   |||
    1044 TGCATCACCTGG...GAGGGCACAAATGGGGAGTTCAAGATGACAGACCC 1090

     284 rAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetA 301
             :::|||||||||   ||||||   ||||||:::::|||   ||||
    1091 TGATGAAGTGGCTCGGCGTTGGGAGAGAGGAAAAGCAAACCTAACATGA 1140

     301 snTyrAspLysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIl  317
         |||||||||||||||||||:::::|||   ||||||   |||:::|||
    1141 ACTATGACAAACTCAGCCGTGCACTTCGCTACTACTATGACAAAAATATT 1190

     318 IleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
         :::   |||         :::::|||   :::|||:::|||
    1191 ATGACTAAAGTTCAT...GGTAAACGCTATGCCTACAAATTT 1229

seq_name: qb ov:AF257741

Seq_documentation_block:
LOCUS       AF257741     1902 bp    mRNA            VRT       24 MAY 2000
DEFINITION  Danio rerio E4tf1 60 transcription factor (e4tf) mRNA, complete
            cds.
ACCESSION   AF257741
VERSION     AF257741.1  GI:8050806
KEYWORDS    .
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopteryqii; Neopteryqii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 1902)
  AUTHORS   Slavov,D.B. and Gardiner,K.
  TITLE     An e4tf cDNA from Zebrafish (Danio rerio)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1902)
  AUTHORS   Slavov,D.B. and Gardiner,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19 APR 2000) Eleanor Roosevelt Institute, 1999 Gaylord
            Street, Denver, Co 80206, USA
FEATURES             Location/Qualifiers
     source          1. .1902
                     /organism "Danio rerio"
                     /db_xref "taxon:7955"
     gene            1. .1902
                     /gene "e4tf"
     CDS             145. .1502
                     /gene "e4tf"
                     /codon_start 1
                     /product "E4tf1 60 transcription factor"
                     /protein_id "AAF71747.1"
                     /db_xref "GI:8050807"
                     /translation "MSKSELEEMIKELHQGKEKQECLEEGIEEQTLIAAELIQQHDI
                     NEPIGNLKKLLEPPIQTPIPCYPPCLQPTQLHPPHSLEPQGVKLQGIVQLSVQYIVKP
                     GEEKLNILEIVKPVETVEVVIDPDAAAGEEAHLVEDGQVIAVERAALPDEISEQVIRW
                     AAALEGYRKEQVRLNIPYDPVQWTADQVIHWAVWVMKEFGIDEMEVGGIHIPGRQLCG
                     FSQFFFLQRVPSGFITWSHLFLIRKYVIASQDQSQEATVTIDQPVQLTPAPVQQATPT
                     AIKVMKHNKIPKAPKISGEERSSIGNKIGNNGQTQEWQFLLELLIDKDSRDCISWVGE
                     EGEFKLNQPELVAQKWGQKKNKPIMNYEKLSRALRYYYDKDMISKVQGKRFVYKFVCD
                     LRTLIGYSAAELNNLVTECEQKKLARVQLQGLGQPVNEVTLATATGQPVTTVTLAAAA
                     LDKDS"
BASE COUNT      501 a    431 c    544 g    426 t
ORIGIN


alignment_scores:
            Quality:  291.50              Length:    284
              Ratio:  1.881                 Gaps:      9
 Percent Similarity:  54.577   Percent Identity: 29.577

alignment_block:
 US-09-126 945B 2 x AF257741   ..

 Align seq 1/1  to: AF257741  from: 1  to: 1902

      95 ValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGl 111
         |||||||||   :::|||   |||||||:::
     498 GTGATCGATCCAGATGCTGCTGCTGGCGAG..................... 527

     111 yLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValG 128
                 |||   |||   :::|||:::   |||   ::::::|||
     528 .........GAAGCTCATCTGGTACAGGATGGACAAGTGATCGCTGTGG 567

     128 lyGluValLeuLysAspIleGluThrAlaCysLysLeu............. 140
         ::::::              |||||||:::   ::::::
     568 AGCGAGCGGCCCTTCCTGATGAGACCTCAGAGCAGGTGACTCGCTGGGCG 617

     141 .....................................LeuAsnIleThrAl 145
                                              |||||||||
     618 GCTGCACTAGAAGGCTACCGCAAGGAGCAGGTTCGACTGAACATCCCTA 667

     145 aAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpT 162
         |||||:::::|||:::   :::::|||      |||   :::|||:
     668 TGACCCCGTGCAGTGGACAGCAGATCAGGTTATCCACTGGGCAGTTTGGG 717

     162 hrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGluLeu 178
         ::      ::::::   :::      |||      :::      ::::::
     718 TGATGAAAGAGTTTGGCATTGATGAAATGGAGGTGGGTGGCATTCACATC 767

     179 AlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSe 195
         |||:::::|||||:::::|||::|||:::|||   ||||||
     768 CCGGGCCGGCAGCTCTGTGGCTTCAGCCAGGAGGAGTTCCTGCAGAGAGT 817

     195 rProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerA 212
         |||   :::|||:::::|||   :::||||||:::::   ::::::
     818 GCCC...AGCGGCGAGATCCTGTGGAGCCACCTGGAGCTTCTGCGCAAGT 864

     212 laAlaTrpMetLysGluArgThrSer..................... 220
         :::      ::::::      |||
     865 ATGTACTTGCGAGTCAGGATCAAAGTCAGGAAGCTACAGTCACCATTGAT 914

     221 ................PrcGlyAlaIleHisTyrCysAlaSerThr.. 230
                         ||||::   ::::::   ::::::   |||
     915 CAACCGGTGCAGATCATCCCTGCTCCTGTGCAGCAGGCCACACCTACAGC 964

     231 ..............................................SerG 232
                                                      |||
     965 CATTAAAGTGATGAAGCACAACAAAACACCACGAGCCCCTCGCATCTCAG 1014
```

```
      232 InGluSerTrpThrAspSerGluValAspSerSerCysSerGlyGlnPro 248
          ||||::  :::::   :::::::::  :::::  :::|||||
     1015 GAGAGGAGCGCACTTCCCCAGGCAACCGCACAGGTAACAACGGTCAA... 1061

      249 IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTy 265
          |||:::|||||||||||||||   ||||||||  |||
     1062 ATTCAGCTGTGGCAGTTTCTCCTGGAGCTGCTGACGGATAAAGACTCTCG 1111

      266 rGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIle 282
          |||   |||::::::  |||:::|||  ||||||:::
     1112 CGAC...TGCATCTCCTGGGTGGGC...GAAGAAGGCGAGTTCAAGCTCA 1155

      283 InAspSerAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgPro 298
          :::::  ||||||::  ||||||  ||||||||||:|||
     1156 ATCAGCCGGAGCTGGTGGCTCAGAAATGGGGTCAGCGCAAGAACAAGCCC 1205

      299 AlaMetAsnTyrAspLysLeuSerArgSerIleArgIleGlnTyrTyrLysGly 315
          :::|||||||||:::|||||||||||||||:::|||  ||||||
     1206 ACCATGAACTACGACAAGCTGAGCAGAGCTCTCAGGATACTACTACGATGGC 1255

      316 sGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnHeV 332
          :::|||  |||  :::  ::::::|||:::|||||:::|||
     1256 AGACATGATCAGCAAAGTGCAA...GGCAAGCGATTGTGTACAAGTTGC 1302

      332 al 332
          ||
     1303 IG 1304

seq_name: gb_pri:HSA276205

seq_documentation_block:
LOCUS       HSA276205     1668 bp    mRNA            PRI       24-NOV-2000
DEFINITION  Homo sapiens mRNA for TELB protein.
ACCESSION   AJ276205
VERSION     AJ276205.1  GI:11342492
KEYWORDS    TELB gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1668)
  AUTHORS   Poirel,H., Lopez,R.G., Lacronique,V., Della Valle,V.,
            Mauchauffe,M., Berger,R., Ghysdael,J. and Bernard,O.A.
  TITLE     Characterization of a novel ETS gene, TELB, encoding a protein
            structurally and functionally related to TEL
  JOURNAL   Oncogene 19 (41), 4802-4806 (2000)
  MEDLINE   20484759
REFERENCE   2  (bases 1 to 1668)
  AUTHORS   Bernard,O.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Bernard O.A., U 444, INSERM - CEPH, 27 rue
            Juliette Dodu Paris, 75010, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1668
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            192. .1217
                     /gene="TELB"
     CDS             192. .1217
                     /gene="TELB"
                     /codon_start=1
                     /product="TELB protein"
                     /protein_id="CAC17012.1"
                     /db_xref="GI:11342493"
                     /translation="MQEGELAISPISPVAAMPPLGTHVQARTEAGENLLGPQQLCKLP
                     GRLRIQPAIWSREDVLHWLRWAEQEYSLPVTAFHGFEMNGRALPTLLKEDTMHRAPSS
                     GDVLYELLQYIKTQKRAIVCGPPEGGIPERLKTPPQHSPVPPPEVTGPSQMDTRRGHLI
                     QPPDPGLTSNFGHLDDPGLARWTPGKEESLNLCHCAELGCRTQGVCSFPAMPQAPIDC
                     RIADCRLLWDYVYQLLLDTRYEPYIEWICESAECFEVVTENGLARWCNRENEVNMTY
                     FKMSKAIRHVYKTNIKKFFGQKLIERRTKTPCKMVQKRSRTEDTSQEQKPEPERTA
                     KRPETSP"
BASE COUNT       391 a    489 c    497 g    291 t
ORIGIN
```

```
alignment scores:
            Quality:  290.50              Length:    407
              Ratio:  1.553                 Gaps:     13
 Percent Similarity:  45.946   Percent Identity: 24.570

alignment block:
 us-09-126-945b-2 x HSA276205   ..

 Align seq 1/1  to: HSA276205  from: 1  to: 1668

       6 ProGlyLeuSerSerValSerProSerHisLeuLeuLeuProProProAspIL 22
         ||||||  :::  ::|||  |||:::      :::|||||
       9 CCGGGGCGCACACGCTTAGCGCTCCTCGAAAGAGATTCCCTCCCACGCG 58

      22 rValSerArgThr.....GlyLeu.....GluLysAlaA 42
         |||  :::          ||||||     |||  ::::
      59 ACCTTCCAGTTCTCGGAGCCAGGTTAGGGGTTGGCGGAGGAGGACTGCG 108

      42 laAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProPro 48
         ::|||  |||  |||                              |||
     109 GGGCGCGGGCCTAGGGC.....................CCC 128

      49 AlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMe 65
         ::::::  :::||||||
     129 AGCAGCCACAGCCAGGG.............................. 146

      66 tLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerS 82
                        |||        :::  |||  |||||||
     147 ..................AGCGCTCAAGACAGAAAGCCGGTGGCTTCCT 177

      82 erArgGluGlnProProGluGluProGluGlnCysProValIleAspSer 98
         ||       ||||||     |||          :::|||:::
     178 CA.....CCTCCACCTGTAATGCAGGAGGAGAATTGGCTATTTCTCCT 221

      99 GlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGl 115
         :::|||:::::::::::::       |||  |||     :::::
     222 ATAAGCCCTCTGCCAGCCATGCCT      CCCCTAGGCACCCACCTGCA 265

     115 uGluHisSerLeuGluGlnValGlnSerMetValValGlyGluValLeuL 132
         :  :::        |||::::::  :::      |||  :::
     266 AGCCAGATGTGAAGCTCAAATTAACCTGCTGGGTGAAGGGGGGATC.... 311

     132 ysAspIleGluThrAlaCysLysLeuLeu...AsnIleThrAlaAspPro 147
                       |||||||||      ::::::      :::|||
     312 ..............TGCAAGCTGCCAGGAAGACTCCGCATCCAGCCC 344

     148 MetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpIleThrGluHi 164
           ||||||  ::::::|||      ||||||  |||:::::::
     345 GCACTGTGGAGCAGGGAGGACGTGCTGCACTGGCTGCTCTGGGTAGAGCA 394

     164 sGlnTyrArgLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyL 181
         :::|||  ||||||     ::::::        |||:::  |||:
     395 GGAGTACTCTCTGCCATGCACCGCGGAGCACGGGTTCGAGATGAACGGAC 444

     181 ysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeu 197
         ::  ||||||  :::::::::::::::::|||||:::|||:::|||
     445 GCGCCTCTGCATCCTCACCAAGGACGACTTCCGGCACGTGCGCCCAGC 494

     198 GlyGlyAspValLeuHis..................................... 203
         :::|||||||||||||:::
     495 TCAGGTGACGTCCTGTATGAGCTGCTCCAGTACATCAAGACCAGCGGCG 544

     204 ..................................................... 203

     545 ACCCCTGGTGTGTGGGCCCTTTTTTGGAGGGATCTTCAGGCTGAAGACGC 594

     203 ..................................                  203

     595 CCACCCAGCACTCTCCAGTCCGCCGGAAGAGGTGACTGGCCCCTCCAG 644

     204 ..................................................... 203
```

```
     645 ATGGACACCCGAAGGGGCCACCTGCTGCAGCCACCAGACCCAGGGCTTAC 694

     204 .........AlaHisLeuAspIleTrpLysSerAlaAlaTrpMetGlySe 237
                 :::||||||||||           |||   |||
     695 CAGCAACTTCGGCCACCTGGATGACCCTGGCCTGGCAAGGTGGACCCCTG 744

     217 luArgThrSerProGlyAlaIleHisTyrCysAlaSer............ 229
         :::   :::         :::   :::||||||:::
     745 GCAAGGAGCACTCCCTCAACTTATGTCACTGTGCACAGCTCGCCTGCAGC 794

     230 ............ThrSerGluGluSerTrpThrAspSerGluValAspSe 262
                     |||       :::       ::::::   :::||||:
     795 ACCCAGCCCGCTCTTCCTTCTTCCCGCAtGCCCCACGGCCCCAtGACGG 844

     242 rSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuL 259
         :      :::         :::   ||||||:::::::   :::||||
     845 CAGGATCGCTGACTGCCGCCTG...CTGTGGGATTACGTGTATCAGCTGC 891

     259 euLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGlu 275
         |||||      :::   |||      ::::||:::|||   :::||:::
     892 TCCTTGATACCCGA...TATGAGCCCTACATCAAGTGGGAAGACAAGGAC 938

     276 LysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGl 292
             ||||||:::::   |||        :::|||||||||||||||
     939 GCCAAGATCTTCGAGTTGTGGATCCAAATGGGCTCGCCAGACTCTGGGG 988

     292 yIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgCys 309
         |   :::||||||||      |||:::|||:::|||:::|||||||:::
     989 AAATCACAAGAACCGGGTGAACATGACCTACGAGAAGATGTCTCGTGCCC 1038

     309 IeArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGln 325
         ::|||:::|||||||||   :::||||||:::|||   :::   :::|||
    1039 TGCGCCACTATTATAAGCTTAATATCATTAAGAAG...GAACCGGGGCAG 1085

     326 ArgLeuValTyrGlnPheVal 332
         :::|||:::::::::|||:::
    1086 AAACTCCTGTTCAGATTTCTA 1106

seq name: qb_ov:DRE249590

seq documentation block:
LOCUS       DRE249590     1436 bp    mRNA            VRT       04 MAR-2000
DEFINITION  Danio rerio mRNA for Fli-1 protein (fli-1 gene).
ACCESSION   AJ249590
VERSION     AJ249590.1  GI:6006476
KEYWORDS    fli 1 gene; Fli 1 protein.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 1436)
  AUTHORS   Brown,L.A., Rodaway,A.R., Schilling,T.F., Jowett,T., Ingham,P.W.,
            Patient,R.K. and Sharrocks,A.D.
  TITLE     Insights into early vasculogenesis revealed by expression of the
            ETS domain transcription factor Fli-1 in wild type and mutant
            zebrafish embryos
  JOURNAL   Mech. Dev. 90 (2), 237-252 (2000)
  MEDLINE   20108585
REFERENCE   2  (bases 1 to 1436)
  AUTHORS   Sharrocks,A.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1999) Sharrocks A.D., School of Biological
            Sciences, University of Manchester, Oxford Road, Manchester, M13,
            9PT, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1..1436
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /dev_stage="20-24hr embryo"
     gene            45..1400
                     /gene="fli-1"
     CDS             45..1400
                     /gene="fli-1"
                     /function="transcription factor"
                     /codon_start=1
                     /evidence=experimental
                     /product="Fli-1 protein"
                     /protein_id="CAB56832.1"
                     /db_xref="GI:6006477"
                     /translation="MeGlIkEALSVVSEDQSLEEEPEYAAAAPLEPKIDMIASGIQDYGQ
                     THKINPIPPQQEWINQPVRVNVKREYDHINGSREESPVIETSVCKCNKMVGGIEASQMNY
                     TGYMDEKCAPPPNMTTNERKVIVPADPSLWSFDHVRQWLDWAIKEYGLQEIDIAMFHS
                     TQKEAIKMSKDDFLELTSVYNIEVLLSHLNYLRESSSSTSYNIESHAIDQSPRTAAKD
                     DASYDAVRRTGWSNNMHSGKGSPTVVSQSVSKNPDQPRPQPDPYQTLGPTSSRLANPG
                     SGQAQLWQFLLELLSDSANAGCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKL
                     SRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPTESIMYKYPSELPYVFSYHA
                     HQQKVNFVSPHPPSMPVTSSNFFGPTTPYWSSPTGGIYPNPSVPRHANSHVPSHLGSY
                     Y"
BASE COUNT       377 a    439 c    349 g    271 t
ORIGIN


alignment scores:
            Quality:  289.50              Length:     392
              Ratio:  1.542                 Gaps:      15
 Percent Similarity:  48.214   Percent Identity: 27.041

alignment block:
 US-09-126-945B-2 x DRE249590   ..

 Align seq 1/1  to: DRE249590  from: 1  to: 1436

       2 GlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLe 18
         |||:::   :::   :::|||||||   ||||||   ::::::   ||||:
      51 GGAACTATTAAGGAGGCGCTGTCAGTGGTGAGTGAAGACCAGTCTCTGTT 100

      18 uProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyA 35
         :   |||      ::::::   :::   |||   ||||:          |
     101 CGAGCCTCCATACGCCGCTGCTGCTCCTTTACCCAAGACAGACATGACTG 150

      35 laValGlyLeuGluArgArgAspTrpSerProSerProProAlaThrPro 51
         ||   |||   :::      ||||::::::   :::            ::|||
     151 CGTCCGGGACTCAG......GATTACGGCCAGACGCACAAAATCAACCCC 194

      52 GluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrPr 68
                                                           ||
     195 ..............................................ATTCC 199

      68 oGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluG 85
         |   :::::::::|||      :::          ::::::   ::::::|
     200 GCCTCACCAACACTCCATTAATCACCCCGTCCGCCTCAACGTCAACCCAC 249

      85 lu..................ProProGluGluProGluGlnCysProVal 95
         ||                  |||:::|||   :::|||   |||
     250 AGTATGACCACACTCAACCGATCCAGACAGTCGCCCGTCGACTGCCACCCTC 299

      96 IleAspSerGlnAlaPr.AlaGlySerLeuAspLeuValProGlyGlyLe 112
                           ||||::   :::   :::   ||||||
     300 ..................GGGAAATGCAATAAAATGGTGGGCGGG.. 326

     112 uThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyG 129
                         |||       |||          ::::||
     327 ..................ACCGAAGCGTCTCAGATGAACTATACTGGCT 357

     129 luValLeuLys.................AspIleGluThrAlaCysLys 139
         :::   :::                 ::::::   |||      :::
     358 ACATGGATGAGAAATGTGCTCCGCCTCCAAACATGACGACCAATGAGAGG 407

     140 LeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGl 156
         :::   :::   |||||||||       ||||||||||::::::|||::
     408 AGAGtcAtcGtCccGcAGACCGGtCtcTGGGtCTCCGGAtcACGTGCG 457

     156 nLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyL 173
         :::|||||   |||:::      :::|||   |||      :::
```

```
     458 GCAGTGGTTAGACTGGGCCATTAAGGAGTACGGTCTTCAGGAGATCGACA 507

     173 ysAla...PheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGlu 188
           |||   |||||||||       ||||||||||||||||   ||||||||:
     508 CGGCCATGTTTCACAGCACAGATGGAAAAGAGCTCTGCAACATGAGCAAA 557

     189 GluGlnPhe...ArgGlnArgSerProLeuGlyGlyAspValLeuHisAl 204
         :::::|||   |||      |||      :::   :::||||||   ::
     558 GACGACTTCCTGAGACTCACCAGCGTTTATAACACCGAGGTCCTTCTCTC 607

     204 aHisGln                                            206
         :|||||:
     608 ACATCTCAATTACCTCAGGGAAAGTAGCTCATCGATATCATACAACACGC 657

     206 ......................                              206

     658 CATCTCACGCTGACCAGTCCCCTCGCCTGGCTGCAAAAGACGAATGCGTC 707

     207 ...AspIleTrpLysSerAlaAlaTrpMetLysGlu.............Ar 218
           |||       :.   ......|||   ......             ..
     708 TATGATGCTGTACGGAGGACGGGCTGGTCGAACAACATCCACAGCGGAAA 757

     219 rThrSerProGlyAlaIleHisTyrCysAlaSerProSerGlyGln     234
         :   ||||||   ::::::          :::|||   :::   :::
     758 AGGCTCTCCAACAGTGGTCTCTCAGAGCGTGTCCAAAAACCCGGATCAGC 807

     234 ..............................SerTrpThrAspSerGlu 249
                                       :::   |||:::||||::
     808 CTCGTCCTCAGCCAGATCCTTATCAGATATTAGGACCCACCAGCAGTCGA 857

     240 ValAspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLy 256
         :::   :::       |||||||||   ||||:|||||||||||||||
     858 CTTGCTAACCCAGGCTCAGGTCAG...ATCCAGCTGTGGCAGTTCCTGCT 904

     256 sGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuA 273
          ||||||  ||            :::   |||      |||   |||   :
     905 GGAGCTGCTGTCCGACAGCGCCAACGCCGGC...TGCATCACCTGG...G 948

     274 snLysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArg 289
         ::      ::|||   ||||||:::   |||      :::|||||||||
     949 AACGAACCAACGCAGAGTTCAAGATCACCGACCCAGACGAGGTGGCACGG 998

     290 leuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSe 326
          ||||||   ||||||::::::|||   ||||||||||||||||||||
     999 CGCTGGGGCGAACGCAAGAGCAAACCCAACATCAACTACGACAAGCTGAG 1048

     306 rArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspI 323
         ||||::::::|||   ||||||   |||:::|||::   |||
    1049 CCGAGCGCTGCGCTACTACTACGACAAGAACATCATGACCAAAGTGCAC. 1097

     324 leSerGlnArgLeuValTyrGlnPhe 331
           :::::|||   :::|||:::|||
    1098 ..GGCAAGCGCTACGCCTACAAGTTC 1121

seq_name: qb_prl0:HUMTF160

seq_documentation_block:
LOCUS       HUMTF160     1781 bp    mRNA            PRI       03-FEB-1999
DEFINITION  Human mRNA for transcription factor, E4TF1-60, complete cds.
ACCESSION   D13318
VERSION     D13318.1  GI:286026
KEYWORDS    E4TF1-60; transcription factor.
SOURCE      Homo sapiens (strain HeLa S3) Hela cell, lambda gt10 library, cDNA
            to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates, Catarrhini, Hominidae, Homo.
REFERENCE   1  (bases 1 to 1781)
  AUTHORS   Watanabe,H.
  JOURNAL   Unpublished (1993)
REFERENCE   2  (sites)
  AUTHORS   Watanabe,H., Sawada,J., Yano,K., Yamaguchi,K., Goto,M. and Handa,H.
  TITLE     cDNA cloning of transcription factor E4TF1 subunits with Ets and
            notch motifs
  JOURNAL   Mol. Cell. Biol. 13 (3), 1485-1491 (1993)
  MEDLINE   93180783
COMMENT     Submitted (30 SEP 1992) to DDBJ by:
            Hiroshi Handa
            Tokyo Institute of Technology
            4259 Nagatsuta-cho, Midori-ku
            Yokohama 227
            Japan
            Phone:  045-922-1111
            Fax:    045-923-0380.
FEATURES             Location/Qualifiers
     source          1. .1781
                     /organism="Homo sapiens"
                     /strain "Hela S3"
                     /db_xref="taxon:9606"
                     /cell_type="Hela cell"
                     /clone_lib="lambda gt10"
     CDS             193. .1557
                     /codon_start=1
                     /product="E4TF1-60"
                     /protein_id="BAA02575.1"
                     /db_xref="GI:286027"
                     /translation="MIKREAEELIEIEILSTEKAECTEESIVEQIYALAECVSQAIDI
                     NFPIGNFKKIFPRIQCSIDAHFICLQDIQLDPPRSLFDQCVKIDGTVQLSVQVISYQ
                     GIEPKLNILEIVKPADTVEVVIDPDAHHAESEAHLVEEAQVITLDGTKHIITISIEIS
                     EQVTEWAAALECYEKEQEELGIPYIPPIQWSTDQVLHWVVWVMKEFSMIDIDLITLNIS
                     GRELCSLNQEDFFQRVPRGEILWSHILELRKYVLASQEQQMNFIVTIDQPVQLIPASV
                     QSAIPTTIKVINSSAKAAKVQPAPPISGFDRSSPGNRTGNNGQIQLWQFLLELLIDKD
                     ARDCISWVGDEGEFKLNQPELVAQKWGQRKNKPTMNYEKLSRALRYYYDGDMICKVQG
                     KEFVYKFVCELKTLIGYSAAELNKLVIECEQKKLAKMQLHGIAQPVIAVALATASLQT
                     EKDN"
BASE COUNT      555 a    355 c    420 g    451 t
ORIGIN


alignment_scores:
            Quality:  289.00              Length:    306
              Ratio:  1.680                 Gaps:      9
 Percent Similarity:  56.209   Percent Identity: 27.778

alignment_block:
 US-09-126-945B-2 x HUMTF160   ..

 Align seq 1/1  to: HUMTF160  from: 1  to: 1781

      60 LeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLy 76
         |||:::   ::::::::::::   |||   |||:::      ::::::
     514 TTAAACATCCTTGAAATTGTTAAACCTGCGGACACTGTTGAGGTTGTTAT 563

      76 sAlaProGlyAlaSerSerArgGluGluProProGluGluProGluGluC 93
             |||   |||         |||:::      :::      |||:::
     564 TGATCCAGATGCCCACCATGCTGAATCAGAAGCACATCTTGTTGAAGAAG 613

      93 ysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValPro 109
         ::   ||||||                  :::||||||   :::
     614 CTCAAGTGATA.....................ACTCTTGATGGCACAAAA 642

     110 GlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetVa 126
                 |||:::::::::::::::::   |||||||||         ::
     643 CACATCACAACCATTTCAGATGAAACTTCAGAACAAGTGACAAGATGGGC 692

     126 lValGlyGluValLeuLysAspIleGlyThrAlaCysLysLeuLeuAsnI 143
         ::::      :::|||:::      :::          :::   |||:::|
     693 TGCT......GCACTGGAAGGCTATAGGAAAGAACAAGAACGCCTTGGGA 736

     143 leThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeu 159
         ||      |||||:::::::|||||   :::::|||      |||:::
     737 TACCCTATGATCCCATACAGTGGTCCACAGACCAAGTCCTGCATTGGGTG 786

     160 LeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGl 176
         :::|||:::      ::::::   :::      :::      :::
     787 GTTTGGGTAATGAAGGAATTCAGCATGACCGATATAGACCTCACCACACT 836
```

```
     176 nGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgG 193
         :::::::::|||:::|||||||||:::::::::::|||:::|||   |
     837 CAACATTTCGGGCAGAGAATTATGTAGTCTCAACCAAGAAGATTTTTTC 886

     194 lnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrp 209
         |||||   |||       |||:::::::|||   :::||||||:::::
     887 AGCGGGTTCCT...CGGGGACAAATTCTCTGGACTCATCTGGAACTTCTC 933

     210 LysSerAlaAlaTrpMetLysGluArq..................... 219
         ::::::   :::       :::::::::
     934 CGAAAATATGTATTGGCAAGTCAAGAACAACAGATGAATGAAATAGTTAC 983

     219 .....................................ThrSerP 221
                                              :::::|
     984 AATTGATCAACCTGTGCAAATTATTCCAGCATCAGTGCAATCTGCTACAC 1033

     221 roGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAsp 237
         ||   ::::|||          |||:::::::::::   :::   ::::::
    1034 CTACTACCATTAAAGTTATAAATAGTAGTGCGAAAGCAGCCAAAGTACAA 1083

     238 ..............SerGluValAspSerSerCysSerGlyGln..... 247
                       |||      |||  |||      ||||::
    1084 AGAGCGCCGAGGATTTCAGGAGAAGATAGAAGCTCACCTGGGAACAGAAC 1133

     248 ............ProIleHisLeuTrpGlnPheLeuLysGluLeuLeuL 260
                     |||:::|||||||||||||||   ||||||||
    1134 AGGAAACAATGCCCAAATCCAACTATGGCAGTTTTGCTAGAACTTCTTA 1183

     260 euLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLys 279
         :::             |||  |||::::::  |||
    1184 CTGATAAGGACGCTCGAGAC...TGCATTTCTTGGGTTGCTGATGAA... 1227

     277 GlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyLl 293
         |||  ||||||::::::::          |||||:::  ||||||
    1228 GGTGAATTTAAGCTAAATCAGCCTGAACTGGTTGCACAGAAATGGGGACA 1277

     293 eArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleA 310
         ||||||||||:::|||:::||||||||||:::||||||||||||:::::|
    1278 GCGTAAAAATAAGCCTACGATGAACTATGACAAACTCAGTCGTGCATTAA 1327

     310 rgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArg 326
         ||  ||||||          :::|||  |||  :::  ::::::|||
    1328 GATATTATTACGATGGGGACATGATTTGTAAAGTTCAA...GGCAAGACA 1374

     327 LeuValTyrGlnPheVal 342
         :::||||||:::||||||
    1375 TTTCTGTACAAGTTTGTC 1392

seq_name: qb_rol:AF049702

seq_documentation_block:
LOCUS       AF049702      2224 bp    mRNA            ROD       27 NOV 2000
DEFINITION  Mus musculus E74-like factor 5 (Ets5) mRNA, complete cds.
ACCESSION   AF049702
VERSION     AF049702.1  GI:3928979
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2224)
  AUTHORS   Zhou,J., Ng,A.Y., Tymms,M.J., Jermiin,L.S., Seth,A.K., Thomas,R.S.
            and Kola,I.
  TITLE     A novel transcription factor, ELF5, belongs to the ELF subfamily of
            ETS genes and maps to human chromosome 11p13-15, a region subject
            to LOH and rearrangement in human carcinoma cell lines
  JOURNAL   Oncogene 17 (21), 2719-2732 (1998)
  MEDLINE   99054671
   PUBMED   9840936
REFERENCE   2  (bases 1 to 2224)
  AUTHORS   Zhou,J., Ng,A.Y.N., Tymms,M.J., Thomas,R.S. and Kola,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-1998) Molecular Genetics and Development Group,
            Institute of Reproduction and Development, Monash Medical Centre,
            246 Clayton Road, Clayton, Victoria 4168, Australia
FEATURES             Location/Qualifiers
     source          1..2224
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            1..2224
                     /gene="ELF5"
     CDS             117..878
                     /gene="ELF5"
                     /note="Ets transcription factor"
                     /codon_start=1
                     /product="E74-like factor 5"
                     /protein_id="AAC79754.1"
                     /db_xref="GI:3928980"
                     /translation="MLDSVTHSTFLPNASFCDPLMPWTDLFSNEDYYPAFEHQTACDS
                     YWTSVHPEYWIKKHVWEWLQFCCDQYKLDANCISFCHPNISGLQLCSMIQEEFTEAAG
                     DGEYLYFIAQNIVSQCYSFFNDAEELRSCTRDDLAQSSCTLRLSCIESGVCTHSPISLGS
                     SHLWEFVKDSLSSPERNCGLEWLDREQCTERVVYESEALAEMWCQPEENDCMIYSSLS
                     KALRYYYKIRTLERVDRKLVYKFGKNAIDWQEEKL"
     polyA_signal    1992..1999
                     /gene="ELF5"
     misc_feature    1721..1725
                     /gene="ELF5"
                     /note="AATTA motif"
     misc_feature    1878..1882
                     /gene="ELF5"
                     /note="AATTA motif"
     misc_feature    2072..2076
                     /gene="ELF5"
                     /note="AATTA motif"
     polyA_signal    2181..2186
                     /gene="ELF5"
     misc_feature    2191..2195
                     /gene="ELF5"
                     /note="AATTA motif"
BASE COUNT      603 a    529 c    481 g    611 t
ORIGIN


alignment_scores:
            Quality:  288.00              Length:     207
              Ratio:  2.215                 Gaps:       5
 Percent Similarity:  62.802   Percent Identity:  62.850

alignment_block:
 03-09-126-945B-2 x AF049702   ..

 Align seq 1/1  to: AF049702  from: 1  to: 2224

     135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
         :::|||||||||                :::::   |||      ||:::
     241 CAGACAGCCTGTGATTCCTACTGGACATCAGTGCACCCTGAATACTGGAC 290

     151 rProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 168
         :      ::::||   :::||||||   :::         ||||||:::|
     291 CAAGCGCCATGTCTGGGAATGGCTCCAATTCTGCTGTGACCACTACAAGC 340

     168 euProProMetGlyLysAlaPhe......GlnGluLeuAlaGlyLysGlu 182
         ||              ::::||          ::::::::|||   :::
     341 TTGATGCCAACTGCATCTCCTTCTGTCACTTCAACATCAGCGGCCTGCAG 390

     183 LeuCysAlaMetSerGluGluGluPheArgGlnArgSerProLeuGlyGl 199
         ||||||:::||::::::|||:::|||   :::   :::   :::   ||
     391 CTCTGCAGCATGAGCAGGAGGAGTTCATTGAGGCAGCCGGCATCTGTGG 440

     199 yAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMet. 215
         |:::   |||:::      |||:::   ::::|||   ::::::
     441 GGAGTACCTGTACTTCATTCTCCAGAACATTCGCTCGCAAGGTTACTCCT 490

     216 .....LysGluArgThrSerProGlyAlaIleHisTyr..... 226
               |||:::|||:::  |||       |||
```

```
     441 TTTTCAATGATGCTGAAGAGACCAAGACTGGCATCAAAGACTATGCTGAT 583

     227 ......CysAlaSerThrSerGluGluSerIrpIhrAspSerGluValAs 241
               |||   :::||||||       :::      :::::::::
     584 TCCAGTTGCTTGAAAACAAGT......GGCATCAAGACTCAAGACTCTCA 671

     241 pSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluL 258
          |||      |||   |||      ||||||||||:::|||:::::::::|
     672 CAGCCGAACAAGCCTCCAAAGTTCTCACCTGTGGGAATTTGTCAGAGACT 624

     258 euLeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLys 274
          |||||||||:::|||::::::      :::::::::|||   ::::::
     625 TGCTGCTGTCCCCTGAAGAGAACTGTGGCATCCTGGAATGGGAAGACAGG 674

     275 GluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTr 291
          |:::::|||||||||:::::      |||      :::|||:::::::||
     675 GAGCAGGGCATTTTCCGAGTGGTTAAGTCACAAGCCCTGGCAAAGATGTG 724

     291 pGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgS 308
          ::::   ||||||:::::      |||:::|||:::|||||||||||:
     725 GGGACAAAGGAAGAAGAATCACAGGATGACGTACGAGAAGCTGAGCCGAC 774

     308 erIleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSer 324
          :::::|||   ||||||||      |||:::::::::   |||
     775 CCCTGAGATACTACTATAAAACGAGAATTCTGGAGCGGGTTGAC...... 818

     325 GlnArgLeuValTyrGlnPhe 331
          :::|||||||||||||:::|||
     819 CGGAGGTTAGTGTACAAATTT 839

Seq_name: db_vi2:REAEV1

seq_documentation_block:
LOCUS       REAEV1          2466 bp    DNA             VRL       16-MAY-1984
DEFINITION  Avian erythroblastosis virus E26 fragment containing a myb/ets
            oncogene sequence.
ACCESSION   X00144
VERSION     X00144.1  GI:61475
KEYWORDS    ets oncogene; myb oncogene; oncogene.
SOURCE      Avian leukemia virus.
  ORGANISM  Avian leukemia virus
            Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
REFERENCE   1  (bases 1 to 2466)
  AUTHORS   Nunn,M.F., Seeburg,P.H., Moscovici,C. and Duesberg,P.H.
  TITLE     Tripartite structure of the avian erythroblastosis virus E26
            transforming gene
  JOURNAL   Nature 306 (5941), 391-395 (1984)
  MEDLINE   84068165
COMMENT     Data kindly reviewed (16 MAY-1984) by M.F. Nunn.
FEATURES             Location/Qualifiers
     source          1..2466
                     /organism="Avian leukemia virus"
                     /strain="E26"
                     /db_xref="taxon:11946"
     CDS             <1..2011
                     /codon_start=2
                     /product="p135 protein"
                     /protein_id="CAA24979.1"
                     /db_xref="GI:809070"
                     /db_xref="SWISS-PROT:P01105"
                     /translation="NSIMPRKVFQPEQYLQFSSKAGLPSATICFQKSSHLMAFAHNPPA
                     GPLPGAGQAPLGSDYPYYHIAEPQNVPQQTPYPVALHVNTVNVPQFAAAATLQKEYNEE
                     DPEKEKRIFELELILMSTFNELFVQQAIPTQNHIIANYPCWHSEIVALCNIMISGDNAPV
                     SCLGEHHHCTFSPPVDHGISEMMSYYMDTIIGSTGPYFLAKPGVMQGASSCPEDPWMP
                     CRLQSACCPPRSCCPPWDPAATQFVPTGLEHYSTDMECALVPLETPSSKFMMSQALKA
                     IFSGFAKFQQKLGTPKDPQQWIETHVPDWVMWAVNFFSLEGVIFQKPVMNGAALCALS
                     KECFLPLAPDFVGDILWEHLELLQKEEAKPYPANGVNAAYPESKYISDYFISYGIEHA
                     QCVPPSEFSFPSFITFSYQTIHPTSSFFILSEKYFNDVSSVLIQPVQLCSTQIDYFI
                     IKQEVVTPDNMCMGKVSRQKIGQDSFFSIFSYDSCDKLIQSWSSQSSFQSTQRVPSY
                     DSFDSFDYPAALPNHKPKGTFKDYVVPDPADMNFEEPVIPAAALACYIGSCPIQLWQFL
                     LELLTDKSCQSFISWTGDGWEFKLSDPDFVARRWGKRKNKPKMDYEKLSRGLRYYYDK
                     NVIHKTAGKPYVYRFVCDIQSLLGYTPEEHSSASGLISSMACSSE"
BASE COUNT      615 a    695 c    657 g    499 t
```

```
ORIGIN


alignment_scores:
            Quality:  288.00              Length:    507
              Ratio:  1.385                 Gaps:     21
 Percent Similarity:  41.026   Percent Identity: 23.866

alignment_block:
 US-09-126-945B-2 x REAEV1    ..

 Align seq 1/1  to: REAEV1  from: 1  to: 2466

      15 HisLeuLeuLeuProProAspThrValSerArgThrGlyLeuGluLysAl 31
         |||                 |||:::::::      :::|||   ::::::||
     428 CACAGCACCACGGTTGCTGACAATACCATGACCAGTGGT...GACAATGC 474

      31 aAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerProP 48
         |   ::::::::::::::|||   |||::::::::      :::|||||||||||
     475 ACCTGTTTCCTGTTTGGGG...GAACATCACCACTGTACTCCATCTCCAC 521

      48 roAlaThrProGluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAsp 64
         ||:::      :::::|||   |||   :::          |||:::|||
     522 CAGTG......GATCATGGCACCTCAGAGATGATGAGTTACTACATGGAC 565

      65 MetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaPro..... 78
                 :::   :::   :::            |||
     566 ACAACCATTGGCAGCACGGGTCCTTATCCTTGGCTCGGCCTGGAGTGAT 615

      79 ....GlyAlaSerSerArgGluGluProProGluGluPro..... 90
             ||:|||||||||      |||  |||    |||
     616 GCAAGGTGCTAGCAGCTGCTGTGAGGACCCCTGGATGCCATGCAGGCTGC 665

      91 ......................GluGlnCysProValIleAspSerGln 99
                                   :::  ||||||     ||||::
     666 AGTCTGCCTGCTGCCCGCCCAGGTCCTGTTGCCCGCCGTGGGATCAGCG 715

     100 Ala............ProAlaGly............SerLeuAspLeuVa 108
         |||           ||:::|||           |||   ||:::
     716 GCCATCCAGGAAGTTCCCACTGGCCTGGAGCACTACAGCACAGACATGGA 765

     109 IProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerM 125
         :::  :::  |||       |||  :::::::::::::|||:
     766 ATGTGCAGATGTGCCTTTGTTAACCCCCAGCAGCAAGGAAATGATGTCTC 815

     125 etValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeu 141
         :::::::     :::::::::  :::          :::   |||
     816 AGGCACTGAAAGCCACCTTCAGTGGCTTCGCAAAGGAGCAGCAGCGGCTG 865

     142 AsnIleThrAlaAspProMetAspTrpSerProSerAsnValGluLysIr 158
         :::|||      |||||||::::::|||:::   ::::::::::::   ||
     866 GGAATCCCCAAAGATCCCCAGCAGTGGACAGACGCGCCGTGCCGGACTG 915

     158 pheuLeuTrpThrGluHisGluTyrArgLeuProProMetGlyLysAlaP 175
         |:::::|||:::   :::::::::   |||         |||       |
     916 GGTGATGTGGGCAGTGAATGAGTTCACGCTG......AAGGGAGTGGATT 959

     175 heGlnGlu......LeuAlaGlyLysGluLeuCysAlaMetSerGluGlu 189
         |||||:::      :::   |||      |||||||||:::::::::|||
     960 TCCAGAAGTTCTGCATGAACGGAGCTGCCCTGTGCGCCCTGGGCAAGGAG 1009

     190 GlnPheArgGlnArgSerPro...LeuGlyGlyAspValLeuHisAlaHi 205
            |||   :::   :::|||   :::   |||||||:::|||      ||
    1010 TGCTTCCTGGAGCTAGCGGCCTGACTTTGTGGGAGAATATCCTTTGGGAACA 1059

     205 stGluAspIleTrpLys..................................... 210
         ||||:::|||   :::
    1060 CCTGGAGATCTTGCAGAAAGAAGAGGCAAAACCATACCCAGCAAATGGAG 1109

     211 ..SerAlaAlaIrpMetLysGlnArg...ThrSerProGlyAlaIleHis 225
           :::|||||||:::   :::::|||   ||||||         |||
    1110 TGAATGCAGCGTATCCAGAATCAAGCTATACTTCAGACTACTTCATTAGT 1159
```

```
  226 Tyr.....CysAlaSerThrSerGlu..... 232
      |||           ||||::    ||||||
 1160 TATGGCATCGAGCACGCACAGTGCGTGCCTCCCTCCGAGTTCTCTGAGCC 1209

  233 .....GluSerTrp..... 235
           ||||||:::
 1210 CAGCTTCATCACAGAGTCCTACCAGACCCTCCATCCCATCAGCTCGGAAG 1259

  235 ..... 235

 1260 AGCTTCTGTCCCTCAACTACGACAACGACTATCCCTCAGTCATCCTTCGT 1309

  236 .....ThrAspSer..... 238
           |||||||||
 1310 GACCCCGTCCAGACGGACTCCCTGCAGACAGACTACTTCACAATCAAGCA 1359

  239 .GluVal.....AspSerSerCysSerGlyGln..... 247
       ||||||       |||:::  |||  |||:::
 1360 AGAAGTGGTAACGCCAGACAACATGTGCATGGGACGTGTCAGTCGAGGTA 1409

  247 ..... 247

 1410 AACTGGGTGGCCAGGACTCCTTTGAGAGCATAGAGAGCTACGACAGCTGT 1459

  247 ..... 247

 1460 GACCGCCTGACACAGTCCTGGAGCAGCCAGTCCTCCTTCCAGAGCTGCA 1509

  247 ..... 247

 1510 GCGCGTCCCCTCCTACGATAGCTTTGACTCAGAGGACTACCCCGCCGCCC 1559

  247 ..... 247

 1560 TGCCCAACCACAAGCCCAAGGGCACCTTCAAGGACTATGTTCGAGATCGG 1609

  247 ..... 247

 1610 GCTGACATGAACAAGGACAAGCCTGTCATTCCTGCCGCTGCCCTCGCCGG 1659

  248 .....ProIleHisLeuTrpGlnPheLeuLysGluLeuL 259
                |||||||::|||||||||||||||||   |||   |
 1660 CTACACAGGCAGTGGACCCATCCAACTGTGGCAATTCCTGCTGGAG...C 1706

  259 euLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGlu 275
      |||||         |||         ||||||   |||   :::   :::
 1707 TGCTCACTGACAAGTCCTGTCAGTCCTTCATCAGCTGGACGGGT...GAT 1753

  276 LysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGl 292
               ||||||::::::|||      :::||||||||   |||||
 1754 GGCTGGGACTTCAAGCTTTCCGATCCAGATGAGGTGGCCAGGCGGTGGGG 1803

  292 yIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerI 309
      |   ||||||||||::|||   |||:::|||:::|||||||||||||:::
 1804 CAAGAGGAAAAACAAGCCCAAGATGGACTATGAGAAGCTGAGCCGTGGCAT 1853

  309 leArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGln 325
      :::||   ||||||   |||::::::|||:::||          ::::::
 1854 TGCGTTACTATTACCACAAGAACGTCATCCACAAG...ACGGCCGGCAAG 1900

  326 ArgLeuValTyrGlnPheVal 332
      |||   ||||||:::||||||
 1901 CGCTACGTCTACCGCTTCGTC 1921

seq_name: qb|em:AF057717

seq_documentation_block:
LOCUS       AF057717      1184 bp    mRNA             MAM       10-JUL-1998
DEFINITION  ovis aries transcription factor GABP alpha subunit mRNA, partial
            cds.
ACCESSION   AF057717
VERSION     AF057717.1  GI:4046689
KEYWORDS    .
SOURCE      sheep.
  ORGANISM  ovis aries
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Ovis.
REFERENCE   1  (bases 1 to 1184)
  AUTHORS   Ezashi,T., Ealy,A.D., Ostrowski,M.C. and Roberts,R.M.
  TITLE     Control of interferon tau gene expression by Ets 2
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (14), 7882-7887 (1998)
  MEDLINE   98318580
REFERENCE   2  (bases 1 to 1184)
  AUTHORS   Ezashi,T. and Roberts,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-1998) Animal Sciences, University of Missouri,
            Columbia, MO 65211, USA
FEATURES             Location/Qualifiers
     source          1..1184
                     /organism="ovis aries"
                     /db_xref="taxon:9940"
                     /dev_stage="conceptus day 13"
     CDS             1..1014
                     /codon_start=3
                     /product="transcription factor GABP alpha subunit"
                     /protein_id="AAC39262.1"
                     /db_xref="GI:4046690"
                     /translation="EVEVVIEPPQAIHHAEAEAHCVEEARVTLLPGLKHIILDSQFVSEQ
                     VIKWAAALEGYRKPQERLGIPYDPIQWSTDQVIHWVVWVMKEFSMIDIHLITLNISGR
                     ELCSLSQEDFFQRVPRGEILWSHLELLRKYVLASQFQQMNEIVTIDQPVQLTPASVQS
                     AFPTAIKAINSSAKAAKVQRAPRISGLPGSSSQNFTGNNGQTQELWQFILELLTEKEAR
                     DCISWVGDEGEFKLNQPEQVAQKWGQREKNEPIMNYEKLSRAIRYYYE*QMLCKVQGER
                     FVYAEVCDLKTLLGYSAAELNRLVIFCEQKKLAKMQLIHGIAQPVTAVALAAASLQTER
                     DI"
BASE COUNT       322 a    231 c    276 g    300 t
ORIGIN


alignment_scores:
            Quality:  283.00              Length:    288
              Ratio:  1.769                 Gaps:      9
 Percent Similarity:  55.556   Percent Identity: 29.167

alignment_block:
 US 09-126-945B-2 x AF057717   ..

 Align Seq 1/1  to: AF057717  from: 1  to: 1184

      78 ProGlyAlaSerSerArgGluGlnProProGluGluProGlnGlnCysPr 94
         |||  |||          |||     :::     ||:::::::
      24 CCAGATCTCACCATCCGGAAGCAGAAGCACATCCTGTCGAGGAACCTCA 73

      94 oValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyG 111
         ||||||                  :::||||||   :::
      74 TGTGATA.....ACTCTTGATGGCACAAAACACA 102

     111 lyLeuThrLeuGlnGluHisSerLeuGlnGlnValGlnSerMetValVal 127
         ||:::::::::::::::::   ||||||||         ::::::
     103 TTACGACCATTTCAGATGAAACGTCAGAACAAGTGACCAGATGGGCGGCT 152

     128 GlyGluValLeuLysAspIleGlnThrAlaCysLysLeuLeuAsnIleTh 144
           :::|||:::     :::         :::   |||:::|||
     153 .....GCACTGGAAGCTTACAGGAAAGAGCAAGAGCGCTCTGGGATACC 196

     144 rAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuI 161
         |||||:::::|||||   :::::|||      |||:::::|
     197 CTATGATCCCATACAGTGGTCCACGCACCAAGTCCTGCATTGGTGGTCT 246

     161 rpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu 177
         |:::      ::::::   :::      :::      :::      :::
     247 GGGTAATGAAGGAATTCAGCATGACTGATATACACCTCACTACACTGAAT 296

     178 LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnAr 194
         ::::::|||:::|||||||||:::::|||:::|||:::|||   |||||
     297 ATTTCTGGAAGAGAACTCTGTAGTCTCAGCCAAGAAGATTTTTTCAGCG 346
```

```
     194 qSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysS 211
         |   |||       |||::::::|||   :::|||||||||::::   :::: 
     847 AGTTCCT...CGGGGAGAAATTCTCTGGAGTCAICTGGAACTTCTCCGAA 893

     211 erAlaAlaTrpMetLysGluArq........................... 218
         ::   :::      :::::::::                           
     894 AATAIGTATTGGCAAGCCAACAACAACAGATGAATGAGATAGTTACAATT 943

     219 .....................................ThrSerProGl 222
                                              :::::|||    
     444 GACCAACCTGTGCAAATTATTCCAGCATCAGTGCAGTCTGCTACACCAAC 493

     222 yAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThrAsp.... 237
          ||||||           |||::::::::::   :::   ::::::    
     494 TGCCATTAAAGCTATAAATAGTAGTGCAAAGGCAGCTAAAGTACAAACAC 543

     238 .....SerGluValAspSerSerCysSerGlyGln.......... 247
                |||     |||   |||     ||:::                
     544 CTCCGAGGATTTCAGGAGAAGATAGAAGTTCGCCTGGAAACAGAACAGGG 593

     248 ProIleHisIleTrpGlnPheIleLysGluLeuLeuGly 268
                 |||:::|||||||||||||||   |||||||||          
     594 AACAATGGCCAGATCCAGCTGTGGCAGTTTTTGCTAGAACTTCTTACTGA 643

     261 sProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyI 279
           :::           |||   |||::::::   |||  |||         
     644 CAAGCATGCTCGAGAC...TGCATTTCTTGGGTTGCTGATGAA...GGAG 687

     278 lePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArg 294
           ||||||:::::::::          ||||||:::   ||||||   |||
     688 AATTTAAGCTAAATCAGCCCGAACTGGTTGCACAGAAATGGGGACAACGT 737

     295 LysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGl 311
         ||||||::|||::|||||||||::|||||||||||||||||::::::||| 
     738 AAAAACAAACCTACAATGAACTATCAGAAACTTAGTCGTGCATTAAGGTA 787

     311 nTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuV 328
          ||||||         :::|||   |||   :::   :::::|||:::| 
     788 TTATTACCATGGGCAGATCATTTGTAAAGTTCAA...GGCAACAGCTTTG 834

     328 alTyrGlnPheVal 332
         |||||::||||||
     835 TCTACAAGTTTGTC 848

seq_name: qb_htq4-AC014838

seq_documentation_block:
LOCUS       AC014838   117376 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014838
VERSION     AC014838.1  GI:6436497
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 117376)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210638 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..117376
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    32921 a  25167 c  25292 g  33996 t
ORIGIN


alignment_scores:
            Quality:  281.00              Length:     337
              Ratio:  1.588                 Gaps:      12
 Percent Similarity:  52.522   Percent Identity: 37.300

alignment_block:
 US-09-126-945B-2 x AC014838/rev  ..

 Align seq 1/1  to reverse of: AC014838  from: 1  to: 117376

      64 AspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAl 80
         |||:::|||   |||   :::::::   ||||||   |||       ::  
   52853 GACGTCCTGAAGCCCACGGAGCCCGCACTGGCTGCCTTGGCCGAGGAGGT 52804

      80 aSerSerArgGluGluProProGlnGluProGluGlnCysProValIleA 97
         :   ::::::   :::||||||||:      :::|||              
   52803 GC::CCCCAGTTCTCC:C:CC:AAACCCCCACCCAC.......      A 52766

      97 spSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThr 113
         |||::::::      :::  :::                  :::       
   52765 AGAGTTCCTGCAGCCAAAGCCCCATCAAAACCCCACTCAAGCCCATCCAC 52716

     114 LeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
         |||:::   :::   |||::|||::::::      |||       ||    
   52715 AAGGAAGACTCCCAAGACGAATCCGTCCAGCGCCAAGCATGTCAACCCAGT 52666

     130 lLeuLys...AspIleGluThrAlaCysLys.............LeuLeuA 142
         ||||:::      :::::::::      |||                 |||:
   52665 CTTGAATTCCGTCCTCCACAGCAAGTTTAACCCAGAACAAATACCCCTGA 52616

     142 snIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrp 158
         ::|||              :::|||:::   ::::::|||       |||  
   52615 ACATACCCGAGCCCCCAACGAATCCACACACCCACACGTGACCTACTCC 52566

     159 LeuLeuTrpThrGlnHisGlnTyrArgLeuProProMetGlyLysAlaPh 175
         |||   ||||::      |||:::::|||      ::::::   :::    
   52565 CTCCAGTCCCCCCTCAAGCAGTTCCAACTAGTTCCCATCAACATCAGCCA 52516

     175 eGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheA 192
         ::::::   |||::|||||||||||||||:::::|||:::|||:::    
   52515 CTCCCAAATGAATCCCAGGAACTCTCCCCCAATCACCCACCAGAGGAATTCA 52466

     192 rgGlnArgSerProLeuGly...GlyAspValLeuHisAlaHisLeuAsp 207
         ::|||:::   |||         |||:::::::::   :::|||||:::  
   52465 ATCAGAAGCTGCCCCCTGATCCAGGCAACATCTTCTGGACCCACCTCCAG 52416

     208 IleTrpLysSerAlaAlaTrpMet.....................LysGl 217
         :::   |||::::::  ::::::                     ::::|| 
   52415 CTCCTAAAACAATCTAATTTCCTCCTCCCTAGTCCACAAACCACCCCACCA 52366

     217 nArg.ThrSerProGlyAlaIleHisTyrCys.................. 227
         |::: :::|||||:::      :::   |||                  
   52365 CCACCCTAACCCCAACCAACCAAGAATCATCCCCCCCAATTCCATATCCA 52316

     227 ..................................................... 227

   52315 CCAATTCAGGGGGCAGCCTTTCCCTTGAACAGCGGATTATGCGGAAATCC 52266

     227 ..................................................... 227

   52265 TATCACTCACTCAACACCTCTCATTGTAAGTACACCTGTCCAATGTTACAC 52216

     228 .........................AlaSerThrSerGluGluSerTrp 245
                                 :::   |||:::   ||||||      
   52215 CAAAGCATCACCTAACAATATCTACTTCGTTCACACCCGTTCAAACTACT 52166

     236 ThrAspSer.....................GluValAspSerSerCysSe 245
         |||:::|||                     :::   |||:::   ::  
```

```
    52165 ACGTCATCCATGAATCCATCAAACTACACTACCATCGGTTCCGGGAATAA 52116

      245 rGlyGluProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysP 262
          :||||||   ::::::|||||||||||||||||   |||   :::|||
    52115 TGGACAA...GTGCAGCTATCGCAGTTGCTACTCGAA...ATACTCACCG 52072

      262 rGlnHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
          ::::::         |||:::|||::::::   |||   |||
    52071 ACTGCGAGCACACGGACGTTATTGAGTGGGTGGGCACCGAG...GGAGAA 52025

      279 PheLysIleGluAspSerAlaGluValAlaArgLeuTrpGlyIleArgLy 295
          |||||:::   |||       :::|||||||||||||||||   :::||
    52024 TTTAAGCTCACTGACCCGGATCGAGTGGCACGCCTTTGGGCCGAGAAGAA 51975

      295 sAsnArgProAlaMetAspTyrAspLysLeuSerAspSerIleArgGlnT 312
          ||||::|||||||||||||||||::|||||||||||||||:::::|||   |
    51974 AAACAAACCCGCTATGAACTACGAAAAGCTATCAAGGGCTCTCCGCTACT 51925

      312 yrTyrLysLysGlyIleIleArgLysProAspIleSerGluArgLeuVal 328
          |||||         ::|||   |||   :::   :::::|||:::::
    51924 ATTACGATGGGCATATGATCTCCAAGCTGTCC...GGCAAGCGATTCGCA 51878

      329 TyrGluPhe 337
          |||:::|||
    51877 TACAAATTT 51869

seq name: qb_inl:AC007984

seq documentation block:
LOCUS       AC007984    174894 bp    DNA                 INV       21-FEB-2001
DEFINITION  Drosophila melanogaster, chromosome 3R, region 97D-97E, BAC clone
            BACR05C10, complete sequence.
ACCESSION   AC007984
VERSION     AC007984.4  GI:13027520
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 174894)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Fartan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garq,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paraqas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phoumanivonq,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zazeri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 3R, region 97D-97E
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174894)
  AUTHORS   Celniker,S.E., Aqbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Fartan,D.E., Gallo,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.P., Kerr,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.K., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Smit,E.,
            Svirskas,P.P., Wan,K.H., Weinburq,T., Zhanq,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 21, 2001 this sequence version replaced gi:6630613.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive web site (http://www.fruitfly.org/Sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1..174894
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
                     /map="97D-97E"
                     /clone="BACR05C10 (D781)"
                     /clone_lib="BPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT     50826 a    36476 c    37056 q    50536 t
ORIGIN


alignment scores:
            Quality:  281.00              Length:     337
              Ratio:  1.588                 Gaps:      12
 Percent Similarity:  52.522   Percent Identity: 27.300

alignment block:
 BS 89 126 945B 2 x AC007984/rev  ..

 Align seq 1/1  to reverse of: AC007984  from: 1  to: 174894

      64 AspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAl 80
         |||:::|||   |||   :::::::::   ||||||   |||       ::
    6125 GACGTCCTGAAGCCCACGGAGGCCGCACTGGCTGCCTTGGCCGAGGAGGT 6076

      80 aSerSerArgGlnGluProProGluGluProGluGlnLysProValIleA 97
         :   ::::::   :::||||||||||      ::::||
    6075 GGTGGGCCAGCTCTCGCGCCGGAAACGGCCAGCCAG.............A 6038

      97 spSerGluAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThr 114
         |||::::::      :::   :::                  :::
    6037 AGACTTCCTCCAGCGAAAGCCCCATCAAAACCCCACTCAAGCGCATGCAC 5988

     114 LeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
         |||:::   :::   |||:::|||:::::      |||      ||
    5987 AAGGAAGACTCCGAAGAGGAATCCGTCGAGGGCAAGGATGTCAACCCAGT 5938

     130 HetLys...AspIleGluThrAlaCysLys.....LeuLeuA 142
         |||:::      :::::::::      |||      |||:
    5937 CTTGAATTGGGTCCTGGACAGCAAGTTTAAGCGAGAACAAATACGCCTGA 5888

     142 snIleThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrp 158
         ::|||                ::|||:::   :::::|||      |||
    5887 ACATACCCGAGGCGGCCAACGAATGCACACGCACGTCGACGTACTGG 5838

     159 LeuLeuTrpThrGluHisGluTyrArgLeuProProMetGlyLysAlaPh 175
         |||   |||:::      |||:::::|||      :::::   :::
    5837 CTGGAGTGGGCCGTCAAGCAGTTCGAACTAGTTGGCATCAACATGAGCGA 5788

     175 eGluGluLeuAlaGlyLysCysLeuCysAlaMetSerGluGlnGlnPheA 192
         :::::   |||:::|||||||||||||||||:::::|||:::|||:
    5787 CTGGCAAATGAATGGCCAGGAACTCTGCGCAATGACGCACGAGCAATTCA 5738

     192 rgGlnArgSerProLeuGly...GlyAspValLeuHisAlaHisLeuAsp 207
         :::|||:::   |||       |||:::::::::   :::||||||:::
    5737 ATCAGAAGCTGCTCCGTGATCCAGGCAACATCTTCTGGACCACTTCCAG 5688

     208 IleIpplysSerAlaAlaIrpMet.....................LysGl 217
         :::   ||||:::::   ::::::                    :::||
    5687 CTGCTAAAAGAATGTAATTTCGTGTCGTACTGCACAAACGAGCGGAGCA 5638

     217 nArgThrSerPro?l,AlaIleHisTyrCys..................... 227
```

```
      |::: :::|||||:::      :::   |||
   5637 GCAGCGIAAGCCCAAGCAACCAAGAATCATGTCGGCCAATTCCATATCGA 5588

    227 ..................................................  227

   5587 CCAATTCAGGGGGCAGCCTTTCGCTTGAACAGCGGATTATCCGGAAATCC 5538

    227 ..................................................  227

   5537 TATCAGTCAGTCAAGAGCTCTGATTGTAAGTAGACGTGTGAAIGTTACAC 5488

    228 ........................AlaSerThrSerGluGluSerTrp  235
                                :::   |||:::   ||||||
   5487 CAAAGCATGAGCTAACAATATCTACTTCGTTCACAGCCGTTGAAAGTACT 5438

    236 ThrAspSer.....................GluValAspSerSerCysSe  245
        |||:::|                      :::   |||:::   ::
   5437 ACGTCATCCATGAATCCATCAAACTACACTACCATCGGTTCCGGGAATAA 5388

    245 rGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysP  262
        :|||||    :::::|||||||||||||||   |||   :::|||
   5387 TGCACAA...GTGCAGCTATGGCACTTGCTACTCCAA...ATACTCACCGA 5344

    263 rGHisSerPyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle  278
          ::::::   |||:::|||::::::   |||   |||
   5343 ACTGCGAGCACACGGACGTTATTGAGTGGGTGGGCACCGAG...GGAGAA 5297

    279 PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLy  295
        |||||:::   |||      :::|||||||||||||||||   :::||
   5296 ITTAAGCTCACTGACCCGGATCGAGTGGCACGCCTTTGGGGGCAGAAGAA 5247

    295 sAsnArgProAlaMetAspTyrAspLysLeuSerArgSerIleArgGlnI  312
        ||||::::|||||||||||||||:::||||||||||||:::::|||   |
   5246 AAACAAACCCGCTATGACTACGAAAAGCTATCAAGCGCTCTCCGCTACTT 5197

    312 yrTyrLysTyrGlyIleIleArgLysProAspIleSerSerArgLeuVal  329
        ||||       :::|||   |||   :::   :::::|||:::::
   5196 ATTACGATGGGGATATGATCTCCAAGGTGTCC...GGCAAGCGATTCGCA 5150

    329 TyrGlnPhe 331
        |||:::|||
   5149 TACAAATTT 5141

seq_name: qb_in2:AE003758

seq_documentation_block:
LOCUS       AE003758   225092 bp    DNA             INV       05-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013286045 section 94
            of 105, complete sequence.
ACCESSION   AE003758 AE002708
VERSION     AE003758.2  GI:10726791
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 225092)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The Genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 225092)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7301501.
FEATURES             Location/Qualifiers
     source          1. .225092
                     /organism "Drosophila melanogaster"
                     /db_xref "taxon:7227"
                     /chromosome="3R"
     mRNA            join(49752. .50996,51114. .54453)
                     /gene "Tl"
                     /note "Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product "CT17414"
                     /db_xref "FLYBASE:FBan0005490"
                     /db_xref "FLYBASE:FBgn0003717"
     gene            <49752. .>54453
                     /gene "Tl"
                     /note "CG5490"
                     /map="97D1-97D2"
                     /db_xref="FLYBASE:FBan0005490"
                     /db_xref "FLYBASE:FBgn0003717"
     CDS             join(49777. .50996,51114. .53187)
                     /gene "Tl"
                     /note "Tl gene product; Nucleotide sequence of the Celera
                     sequence differs from the published sequence for this
                     transcript"
                     /codon_start 1
                     /db_xref "FLYBASE:FBan0005490"
                     /db_xref "FLYBASE:FBgn0003717"
                     /protein_id "AAF56624.1"
                     /db_xref "GI:7301502"
                     /translation="MSELEAASFLALIVTTLQLQWFSSEASFGKDACSEMSIDGLCQ
                     CAEIMSEYEIFCPANAENPTEELIQPEDYVQIMCNEIGEEYQSDKLEIGEVDRV
                     QMRRCMLPSDEIPIASFILDYLGIVSPTLLIPESNNLGMNITPQHIDPIHICKEFRFCIR
                     KLIRIFANLLIGMKNLSHLLELKANIEEMFSHLFDVLENLESILECCNGLKQMPRGIFG
                     KMPELKQLNLWSNQLHNLISKDFFGATSVLGIDIHDNGIEQLPHDVFAHLINVIDINL
                     SANLFRSLPQGLFDHNKHLNEVRLMNNRVPLATIPSRIFANQPFLQTIRLRAFLQSLP
                     GLGLFEHLFQLTNLELSGNKLKTLFATIPQVNTLSLSGNNGLTHLTSSLFAHLINL
                     IDLRLEDNLLTGISGDIFSNLGNLVTLVMSRNRLKTIDSKAFVSTNGLRHLHLDHNEI
                     PLQQFLLLTIMLQTGINGFTVYMHGLLFLNLRNNSITFVYNDWKNIMQIRELILSYNN
                     ISSIGYEDLAFLSQNRLHVNMTHNKIRRIALPEDVHLGEGYNNNLVHVDLNDNPLVCD
                     CTILWFIQLVPGVHKPQYSPQEFKLRTIDRLVCSQPNVLFGTPVPQTEPQTLICPLIFSD
                     PPEERCPEGCNCHVSTVDKALVINCHSQNLTHVPPLPNLHKNMQLMELHLENNLLER
                     LPSANFPGYTSVISLHLAGNNLTSIDVHQLPTNLIHLDISWNHLQMLNAIVIGFLNRL
```

```
                     MFWPSVELSSLNFWMPSFSAFFLLFIQENFLFSCHNRGMSVNAAMPSHMVFLSFNDI
                     CPAEKCVPIALAVVIALPSLLAGPIAALYFFFQTFFEPWLVIGCFIPFIFFLFFEI
                     KKFDAFISYSHKDQSFTFDYLVPQLEHGPQKFQLCVHERDFWLVCGHIPENIMRSVAQS
                     RRTIIVLSQNFTKSEWAKLEFRAAHKSALNEGRSRLIVIFYSQVDGAVEKLDEELKAYL
                     EMNTYLKWGDFWFWVQKLAIALPHKFFVGNIGNGALECALKLLLLKICIVIF
                     PLTPPPAEATENPLVAQLNFVTPHQATMIANGENPCLNLYLPNFGLIGNCRNCAFTT
                     NTNAKQSDV"
mRNA                 complement(join(<55222..55302,55485..55615,55695..55966,
                     56067..>56135))
                     /gene="CG14255"
                     /product="CT33875"
                     /db_xref="FLYBASE:FBgn0014255"
                     /db_xref="FLYBASE:FBqn0034455"
                     /evidence=not_experimental
gene                 complement(<55222..>56135)
                     /gene="CG14255"
                     /map="97D1-97D2"
                     /db_xref="FLYBASE:FBgn0014255"
                     /db_xref="FLYBASE:FBqn0034455"
                     /evidence=not_experimental
CDS                  complement(join(55222..55302,55485..55615,55695..55966,
                     56067..56135))
                     /gene="CG14255"
                     /note="CG14255 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBgn0014255"
                     /db_xref="FLYBASE:FBpp0089459"
                     /evidence=not_experimental
                     /protein_id="AAF56625.1"
                     /db_xref="GI:7301503"
                     /translation="MPAITPSSKKQQSSSSATNLRKCPVEAQCNLLYVQFTPSIGFFLSL
                     GAIQQNFESNGSSNSSSSSSSSTSTTSSTTQASPACPLTRAATSTETTVCPPAAITP
                     IGPTASVGTTIGATLSVTFCVTCIGLLAFSPARRQRTRRLFRRSNSAVRYSRVQSNEE
                     ANLLLEPNGEFTESDDDMLL"
mRNA                 complement(join(<60654..60827,60891..61070,61443..61615,
                     61704..>61779))
                     /gene="CG14256"
                     /product="CT33876"
                     /db_xref="FLYBASE:FBgn0014256"
                     /db_xref="FLYBASE:FBqn0039456"
                     /evidence=not_experimental
gene                 complement(<60654..>61779)
                     /gene="CG14256"
                     /map="97D1-97D2"
                     /db_xref="FLYBASE:FBgn0014256"
                     /db_xref="FLYBASE:FBqn0039456"
                     /evidence=not_experimental
CDS                  complement(join(60654..60827,60891..61070,61443..61615,
                     61704..61779))
                     /gene="CG14256"
                     /note="CG14256 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBgn0014256"
                     /db_xref="FLYBASE:FBqn0039456"
                     /evidence=not_experimental
                     /protein_id="AAF56626.1"
                     /db_xref="GI:7301504"
                     /translation="MEVCLKIQENESLNKCKPSCAGLIFLFSLLLSSPSSTHGFFLVSLI
                     QRNRTFAELAYLKCDPCPTDNTSELSTHLLFNCNHRACPSQPVLRCVEECAYRFTWET
                     NVFCPPHFCTFSADFCDLVHDELCGKEFNFKSACFLSSKLLCLCQNATEMSVNICGAHR
                     KAMLDYSQALVNLFFTHESPNCGKEGEFFCVHWVLFMQSG"
mRNA                 complement(join(<67326..67494,67613..67726,67883..68337,
                     68442..>68454))
                     /gene="CG6396"
                     /product="CT19972"
                     /db_xref="FLYBASE:FBgn0000396"
                     /db_xref="FLYBASE:FBqn0039457"
gene                 complement(<67326..>68454)
                     /gene="CG6396"
                     /map="97D2-97D2"
                     /db_xref="FLYBASE:FBgn0000396"
                     /db_xref="FLYBASE:FBqn0039457"
CDS                  complement(join(67326..67494,67613..67726,67883..68337,
                     68442..68454))
                     /gene="CG6396"
                     /note="CG6396 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBgn0000396"
                     /db_xref="FLYBASE:FBqn0039457"
                     /protein_id="AAF56627.1"
                     /db_xref="GI:7301505"
                     /translation="MLVFCACYEYNGTLTGLSNCNCQLRFNYTGSEYTLTYEVCAFCSI
                     AIGMRKWSTKTLVCANNATKSNGACIAVSVVCLKTLESNCQLLQYQTPLACKEFI
                     FCKATTYVDEFNFSCSSGSELHEFCLPATLNYEAPVCLPACMCPLVPKTLKFPLN
                     VCPCLVPEYQLGYAGSSAACMAKVIAAGAPEEEPSCAYAYEYGPFGWGGMGVLAI"
mRNA                 complement(join(72585..72607,72884..72996,73062..73183,
                     73241..73388,73448..74721,74810..74973,75031..75245,
                     74509..74606,74677..74750,75091..>75141))
                     /gene="CG6390"
                     /product="CT19946"
                     /db_xref="FLYBASE:FBgn0000390"
                     /db_xref="FLYBASE:FBqn0039458"
gene                 complement(<72585..>75141)
                     /gene="CG6390"
                     /map="97D2-97D2"
                     /db_xref="FLYBASE:FBgn0000390"
                     /db_xref="FLYBASE:FBqn0039458"
CDS                  complement(join(72585..72607,72884..72996,73062..73183,
                     73241..73388,73448..74721,74810..74973,75031..75245,
                     74509..74606,74677..74750))

alignment_scores:
            Quality:  281.00              Length:    337
              Ratio:  1.588                 Gaps:     12
 Percent Similarity:  92.522   Percent Identity:  87.300

alignment_block:
 US-09-126-945B-2 x AE003758/rev   ..

 Align_seq 1/1  to reverse of: AE003758  from: 1  to: 225092

      64 AspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAl 80
         |||:::|||  |||  :::::::::  ||||||  |||      ::
  124256 GACCTCTACTGAAGCCCACCGACGCCCCACTCGCTGCCTTCGCCCCAGCGCC 124207

      80 aSerSerArgGluGlaProProGluProGluGluCysProValIleA 97
         :  ::::::  :::|||||||||      :::|||
  124206 GGTGGGCAGCTCTCGCGCGCCGGAAACCGGCAGCCAG.............A 124169

      97 spSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThr 113
         |||::::::      :::   :::          :::
  124168 AGACTTCCTCCACGCAAAGCCCCATCAAAACCCCACTCAAGCGCATGCAC 124119

     114 LeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
         ||:::   :::   |||:::|||:::::      |||      ||
  124118 AAGGAAGACTCCGAAGAGGAATCCGTCGAGGGCAAGGATGTCAAGCCAGT 124069

     130 IleuLys...AspIleGlnIleAlaCysLys.............LeuLeuA 142
         |||:::      :::::::::      |||                 |||:
  124068 CTTGAATTGGGTCCTGGACGCAAGTTTAAGCGAGAACAAATACGCCTGA 124019

     142 spIleIleAlaAspProMetAspTrpSerProSerAsnValGluLysTrp 158
         ::|||          :::||:::   :::::|||     |||
  124018 AGATACCCGAGGCGCCAACGAATGGACACACGCACGTGCAGCTACTGG 123969

     159 LeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPh 175
         |||   |||:::      |||:::::|||      :::   :::
  123968 CTGGATTGGCTCTCAAGCAGTTCGAACTATTGGCATCAACAAGCAGCGA 123919

     175 euIleGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGluPheA 192
         ::::::   |||:::|||||||||||||||:::::|||:::|||
  123918 CTGGCAAATGAATGCCAGGAACTCGCGCAATGACGCACGAGGAATTCA 123869

     192 rgGlnArgSerProLeuGly...GlyAspValIleuHisAlaHisLeuAsp 207
         :|||:::  |||      |||:::::::::  :::||||||:::
  123868 ATCAGAACCTGCCCCGTGATCCAGGCAACATCTTCTGGACCCACCTCCAG 123819
```

```
    208 IleIrpLysSerAlaAlaIrpMet.....LysGl 217
        :::  ||||||||  ::::::      :::||
 123818 CTGCTAAAAGAATGTAATTTCGTGTCCGTAGTGCACAAACGAGCGGACGCA 123869

    217 qArg.ThrSerProGlyAlaIleHisTyrCys..... 227
        |:::.:::|||||||:::      :::  |||
 123768 GCAGCGTAAGCCCAAGCAACCAAGAATCATGTCGGCCAATTCCATATCGA 123719

    227 ..... 227

 123718 CCAATTCAGGGGGCAGCCTTTCGCTTCAACAGCGGATTATAGCGGAAATCC 123669

    227 ..... 227

 123668 TATCAGTCAGTCAAGAGCTCTGATTGTAAGTAGACGTGTGAATGTTACAC 123619

    228 .....AlaSerThrSerGluGluSerTrp 235
              :::  |||:::  ||||||
 123618 CAAAGCATGAGCTAACAATATCTACTTCGTTCACAGCGGTTGAAAGTACT 123569

    236 ThrAspSer.....GluValAspSerserTyrse 247
        |:::::||          :::  ||||::  ::
 123568 ACGTCATCCATGAATCCATCAAACTACACTACCATCGGTTCCGGGAATAA 123519

    245 rGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysP 262
        :|||||||  ::::::||||||||||||||||  |||  :::|||
 123518 TGGACAA...GTGCAGCTATGGCAGTTCCTACTCGAA...ATACTCACCG 123475

    262 roHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
        ::::::        ||||:|||||||:  |||  |||
 123474 ACTGCGACCACACGGACGTTATTGAGTGGGTGGGCACCGAG...GGAGAA 123428

    279 PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLy 295
        ||||||||  |||      :::|||||||||||||||||  ::|||
 123427 TTTAAGCTCACTGACCCGGATCGAGTGGCACGCCTTGGGGGCGACAAGAA 123378

    296 sAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnP 312
        |:::::||||||||||||||||||::||||||||||||::::::|||  |
 123377 AAACAAACCCGCTATGAACTACGAAAAGCTATCAAGGGCTCTCCGCTACT 123328

    312 yrTyrLysLysGlyIleIleArgLysProAspIleSerGluArgLeuVal 328
        |||||        :::|||  |||  :::  ::::::||||:::
 123327 ATTACGATCGGGATATGATCTCCAAGGTGTCC...GGCAAGCGATTCGCA 123281

    329 TyrGlnPhe 331
        :|:::::|||
 123280 TACAAATTT 123272

seq_name: qb_pr5:AF116509

Seq_documentation_block:
LOCUS       AF116509     1551 bp    mRNA            PRI       19-MAR-2001
DEFINITION  Homo sapiens Ets transcription factor TEL-2b (TEL-2) mRNA, complete
            cds.
ACCESSION   AF116509
VERSION     AF116509.1  GI:5442097
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1551)
  AUTHORS   Gu,X., Shin,B.H., Akbarali,Y., Weiss,A., Boltax,J., Oettgen,P. and
            Libermann,T.A.
  TITLE     Tel-2 is a novel transcriptional repressor related to the ets
            factor tel/etv-6
  JOURNAL   J. Biol. Chem. 276 (12), 9421-9426 (2001)
  MEDLINE   21154798
REFERENCE   2  (bases 1 to 1551)
  AUTHORS   Shin,B.H., Boltax,J., Gu,X., Akbarali,Y. and Libermann,T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1551
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21"
                     /tissue_type="pancreas"
     gene            1..1551
                     /gene="TEL-2"
     CDS             76..1101
                     /gene="TEL-2"
                     /codon_start=1
                     /product="Ets transcription factor TEL-2b"
                     /protein_id="AAD43251.1"
                     /db_xref="GI:5442098"
                     /translation="MQEGELAISPISPVAAMPPLGTHVQARCEAQINLLGEGGICKLP
                     GRLRIQPALWSSEDVILHWLRWAEQEYSLPCTAEHGFEMNGRALCLLTKDDFRHRAFSS
                     GDVLYELLQYIKTQRRALVCGPFFGGIERLKTPTQHSPVPPEEVTGPSQMDIRRGHLL
                     QPPDPGLTSNFGHLDDPGLARWTPGKEESLNLCHCAELGCRTQGVCSFPAMPQAPIDG
                     RIADCRLLWDYVYQLLLLDTRYEPYIKWEDKDAKIFEVVDPNGLARLWGNHKNRVNMIY
                     EKMSRALRHIYERLNIIRPEIPCGKELLEFFELKIPGFMVQGKRSHLEPLESQEQDRIFFKI
                     KRPEISP"
BASE COUNT      371 a    455 c    454 g    271 t
ORIGIN


alignment_scores:
            Quality:  280.50              Length:     430
              Ratio:  1.753                 Gaps:       9
 Percent Similarity:  48.485   Percent Identity: 25.152

alignment_block:
 US 09 126 945B 2 x AF116509   ..

 Align seq 1/1  to: AF116509  from: 1  to: 1551

      76 LysAlaProGlyAlaSerSerArgGluGluProProGluGluProGlyGl 92
         :::   |||   |||||||||          ||||||      |||
      43 AGAAAGCCGGGTGGCTTCCTCA       CCTCCACCTGTAATGCAGGAGGG 86

      92 nCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValP 109
                ::|||:::      :::|||:::::::::::::          |
      87 AGAATTGGCTATTTCTCCTATAAGCCCTGTGGCAGCCATGCCT......C 130

     109 roGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMet 125
         ||  |||      ::::::   :::         |||::::::   :::
     131 CCCTAGGCACCCACGTGCAAGCCAGATGTGAAGCTCAAATTAACCTGCTG 180

     126 ValValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeu.. 141
             |||   :::                     |||||||||
     181 GGTGAAGGGGGGATC.....................TGCAAGCTGCCAGG 209

     142  AsnIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysT 158
          ::::::      :::|||      ||||||   ::::::|||      |
     210 AACACTCCGCATCCAGCCCGCACTGTGGAGCAGGGAGGACGTGCTGCACT 259

     158 rpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAla 174
         |||||   |||:::|||:::::::|||   ||||||      ::::::
     260 GGCTGCGCTGGGCAGAGCAGGAGTACTCTCTGCCATGCACCGCGGAGCAC 309

     175 PheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPh 191
         ||::::  ||::::  |||||  ::::::::::::::::||
     310 GGGTTCGAGATGAACGGAGCGCGCCTCTGCATCTTCACCAAGGAGGACTT 359

     191 eArgGlnArgSerProLeuGlyGlyAspValLeuHis..... 203
         ||||:::|||:::|||  :::|||||||||||||:::
     360 CGGGCACGTCTCGCCCAGCTCAGTGACCTCCTGTATGAGCTGCTGCAGT 409

     203 ..... 203

     410 ACATCAACACCCAGCGGCCAGCCCTGGTGTGTGGGCCCTTTTTTGGAGGG 459
```

```
     203 ..................................................  203

     460 ATCTTCAGGCTGAAGACGCCCACCCAGCACTCTCCAGTCCCCCCGGAAGA  509

     203 ..................................................  203

     510 GGTGACTGGCCCCTCTCCAGATGGACACCCGAAGGGGCCACCTGCTGCAGC  559

     204 ............................AlaHisLeuAspIleTrpLys  220
                                    :::|||||||||
     560 CACCAGACCCAGGGCTTACCAGCAACTTCGGCCACCTGGATGACCCTGGC  609

     221 SerAlaAlaTrpMetLysGluArgThrSerProGlyAlaIleHisTyrCy  227
         |||  |||          :::  :::          :::   :::||
     610 CTGGCAAGGTGGACCCCTGGCAAGGAGGACGTCCCTCAACTTATGTCACTG  659

     227 sAlaSer.......................ThrSerGluGluSerTrpI  236
         ||||::                     |||       :::
     660 TGCAGAGCTCGGCTGCAGGACCCAGGGGGTCTGTTCCTTCCCCCGGATGC  709

     236 hrAspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrp  252
         ::::::   :::|||:::       :::         :::   ||||||
     710 CGCAGGCCCCCATTGACGGCAGGATCGCTGACTGCCGCCTG...CTGTGG  756

     253 GlnPheLeuLysGluLeuLeuLysProHisSerTyrGlyArgPheIl  269
         :::::::::   :::||||||||||      :::   |||       ::||
     757 GATTACGTCTATCAGCTGCTCCTTGATACCCGA...TATGACCCCTACAT  803

     269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaG  286
         |:::|||   :::||||::          ||||||||||:   |||
     804 CAACTGGGAAGACAACGACGCCAAGATCTTCGACTTGTGGATCCAAATG  853

     286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr  302
         :::||||||||||||||||   :::||||||||||       |||:::|||
     854 GGCTCGCAGACTCTGGGCAAAATCACAAGAACCGGGTGAACATGACCTAC  903

     303 AspLysLeuSerArgSerIleArgIleTyrTyrLysLysGlyIleIleAr  319
         :::|||:::||||||:::::|||:::||||||||||   :::||||||::
     904 GAGAAGATGTCTCGTGCCCTGCGCCACTATTATAAGCTTAATACATCATAA  953

     319 gLysProAspIleSerGlnArgLeuValTyrGlnPheVal  332
         :|||   :::   :::|||:::|||::::::::::|||:::
     954 GAAG...GAACCGGGGCAGAAACTCCTGTTCAGATTTCTA  990

seq_name: qb_pri0:S45205

seq_documentation_block:
LOCUS       S45205       1673 bp    mRNA            PRI       08-MAY-1993
DEFINITION  Fli-1 Friend leukemia integration 1 [human, mRNA, 1673 nt].
ACCESSION   S45205
VERSION     S45205.1  GI:257454
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1673)
  AUTHORS   Prasad,D.D., Rao,V.N. and Reddy,E.S.
  TITLE     Structure and expression of human Fli-1 gene
  JOURNAL   Cancer Res. 52 (20), 5833-5837 (1992)
  MEDLINE   93007076
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 115336] from the original journal article.
            This sequence comes from Fig. 1A.
FEATURES             Location/Qualifiers
     source          1..1673
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            134..1492
                     /gene="Fli-1"
                     /note="Friend leukemia integration 1"
     CDS             134..1492
                     /gene="Fli-1"
                     /note="Friend leukemia integration 1; This sequence comes
                     from Fig. 1A"
                     /codon_start=1
                     /protein_id="AAB24637.1"
                     /db_xref="GI:257454"
                     /translation="MDGLIKEALSVVSDDQSLFDSAYGAAAHLPKADMTASGSPDYGQ
                     PHKINPLPPQQEWINQPVRKVNVKREYDHMNGSRESPVDCSVSKCSKLVGGGESNPMNY
                     NSYMDEKNGPPPPNMTTNERRVIVPADPTLWTQEHVRQWLEWAIKEYSLMEIDTSEFQ
                     NMDGKELCKMNKEDFLRATFLYNTEVLSHLSYLRESSLLAYNTTSHTDQSSRLSVEE
                     DPSYDSVRRGAWGNNMNSGLNKSPFLGQAQTLSKNTLQQPQPDPYQILGPTSSRLANP
                     GSGQIQLWQFLLELLSDSANASCITWEGTNGEFKMTDPDEVARRWGQRKSKPNMNYDK
                     LSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPTESSMYKYPSDISYMPSYH
                     AHQQKVNFVPPHPSSMPVTSSSFFGAASQYWTSPTGGIYPNPNVPRHPNTHVPSHLGS
                     YY"
BASE COUNT      428 a    488 c    429 g    328 t
ORIGIN


alignment scores:
            Quality:  278.50            Length:    382
              Ratio:  1.522               Gaps:     18
 Percent Similarity:  47.906   Percent Identity: 28.010

alignment block:
 DS:09-126-945b-2 x S45205  ..

 Align seq 1/1  to: S45205  from: 1  to: 1673

      34 GlyAlaValGlyLeuGluArg.........................Ar  41
         |||::|||||   ||||||                            ||
     154 GGCTCTGTCGGTGGTGAGCGACGACCAGTCCCTCTTGACTCAGCGTACG  203

      41 gAspTrpSerProSerProProAlaThrProGluGluGlyLeuSerAlaP  58
         |:::   ||||||||||||                 ||||||       :::
     204 GAGCGGCAGCCCATCTCCC.............CAAGGC.......... 228

      58 hrTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAla  74
                         |||   |||
     229 ..........CGACATGACTGCCTC.......................... 244

      75 AlaLysAlaProGlyAlaSer...SerArgGluGluProProGluGluPr  90
                    |||   |||       |||       |||   :::::::||
     244 ............GGGAGTCCTGACTACGGGCAGCCCCACAAGATCAACCC 281

      90 oGluGlnCysProVal...IleAspSerGlnAlaProAlaGlySer.... 104
         (            :::   :::|||:::::::|||      ||||::
     282 CCCTCCCACCACAGCAGGAGTGGATCAATCAGCCAGTGAGGGTCAACCTC 331

     105 .........................LeuAspLeuValProGlyGlyLeu 112
                                  ::::::   |||   |||||||||
     332 AAGCGGGAGTATGACCACATGAATGGATCCAGGGAGTCTCCGGTGGACTG 381

     113 .....................................ThrLeuGluGlnHi 117
                                          ::::::   ||
     382 CAGCGTTAGCAAATGCAGCAAGCTGGTGGGCGGAGGCGAGTCCAACCCCA 431

     117 sSerLeuGluGluValGluSerMetValVal.GlyGlnValLeuLysAsp 133
         |:::|||:::::::::   :::          |||          :::
     432 TGAACTACAACAGCTATATGGACGAGAAGAATGGCCCCCCTCCTCCCAAC 481

     134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTr 150
         :::   |||      :::   :::   :::   |||||||||      ||
     482 ATGACCACCAACGAGAGGAGAGTCATCGTCCCCGCAGACCCCACACTGTG 531

     150 pSerProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlaTyrA 167
         |:::   :::::::|||:::::|||||   |||:::       :::|||
     532 GACACAGGAGCATGTGGAGGCAATGGCTGGAGTGCGCCATAAAGGAGTACA 581

     167 rgLeuProProMetGlyLysAla...PheGlnGluLeuAlaGlyLysGlu 182
           |||      :::      :::   ||||||:::::   ||||||||
     582 GCTTGATGGAGATCGACACATCCTTTTTCCAGAACATGGATGGCAAGGAA 631

     183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeu...Gl 198
```

```
      |||||||   |||:::::::|||:::|||   :::   :::   |||   ::
    632 CTGTGTAAAATGAACAAGGAGGACTTCCTCCGCGCCACCACCCTCTACAA 681

    198 yGlyAspValLeuHisAlaHisLeu..................... 206
        :   ::::|||||||   :::|||||
    682 CACGGAAGTGCTGTTGTCACACCTCAGTTACCTCAGGGAAAGTTCACTGT 731

    206 ..................................................... 206

    732 TGGCCTATAATACAACCTCCCACACCGACCAATCCTCACGATTGAGTGTC 781

    207 .................AspIleTrpLysSerAlaAlaTrpMetLysGl 217
                        |||      :::   :::|||||||   :::::
    782 AAAGAAGACCCTTGTTATGACTTAGTCAGAAGAGGAGCTTGGGGCAATAA 831

    217 qArgThrSerPro.....................GlyAlaIleHisTyrC 227
        :   :::|||                       ||||::   :::
    832 CATGAATTCTGGCCTCAACAAAAGTCCTCCCCTTGGAGGGGCACAAACCA 881

    227 ysAlaSerThrSerGluGlu............................. 238
          ::::::::::::::|||::
    882 TCAGTAAGAATACAGAGCAACGGCCCCAGCCAGATCCGTATCAGATCCTC 931

    238 SerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHi 258
        :::   |||:::|||||||:   :::       |||||||||   |||::
    932 GGCCCGACCAGCAGTCGCCTAGCCAACCCTGGAAGGGGCAG...ATCCA 978

    259 sLeuTrpGlnPheLeuLysGlyLeuLeuLeuLysProHisSerTyrGlyA 287
        ||||||||||||||||   |||   ||||||:::       :::   :::
    979 GCTGTGGCAATTCCTCCTGGAG...CTGCTCTCCGACAGCGGCCAACGGCA 1028

    287 rgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAsp 306
          |||   |||   :::       :::|||   ||:|||:::   :::
   1029 GCTGTATCACCTGG...GAGGGGACCAACGGGGAGTTCAAAATCACGGAC 1072

    284 SerAlaGluValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMe 303
          :::|||||||||   ||||||   |||||||:::::|||   ||
   1073 CCCGATGAGGTGGCCAGGCGCTGGGGGCAGCGGAAAAGCAAGCCCAACAT 1122

    300 tAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyI 317
        ||||||||:|||||||||||||||:::::::||   ||||||   |||:::|
   1123 GAATTACGACAAGCTGAGCCGGGCCCTCCGTATTACTATGATAAAAACA 1172

    317 leIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
        |:::   |||         :::::|||   :::|||:::|||
   1173 TTATGACCAAAGTGCAC...GGCAAAAGATATGCTTACAAATTT 1213

seq_name: qb_pat1:A36461

seq_documentation_block:
LOCUS       A36461        2916 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 2 from Patent WO9323549.
ACCESSION   A36461
VERSION     A36461.1  GI:2293779
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2916)
  AUTHORS   Aurias,A., Delattre,O., Desmaze,C., Melot,T., Peter,M.,
            Plomastel,B., Thomas,G. and Zucman,J.
  TITLE     NUCLEIC ACID CORRESPONDING TO A GENE OF CHROMOSOME 22 INVOLVED IN
            RECURRENT CHROMOSOMAL TRANSLOCATIONS ASSOCIATED WITH THE
            DEVELOPMENT OF CANCEROUS TUMORS
  JOURNAL   Patent: WO 9323549-A 2 25-NOV-1993;
            CENTRE NAT RECH SCIENT (FR)
COMMENT     Other publication FR 2691475 931126.
            Other publication JP 8500964T 960206.
FEATURES             Location/Qualifiers
     source          1..2916
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /dev_stage="ADULTE"
BASE COUNT      844 a    686 c    663 g    724 t
ORIGIN


alignment_scores:
            Quality:  278.50              Length:    382
              Ratio:  1.522                 Gaps:     18
 Percent Similarity:  47.906   Percent Identity: 28.010

alignment_block:
 US-09-126-945B-2 x A36461   ..

 Align seq 1/1  to: A36461  from: 1  to: 2916

     34 GlyAlaValGlyLeuGluArg.....................Ar 41
        |||::::::::   ||||||                       ||
    141 GGCTCTGTCGGTGGTGAGCGACGACCACTGCCCTCTTTGACTCAGGTACG 190

     41 qAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerAlaP 58
        |:::   ||||||||:|||                 ||||||
    191 GAGCGGCAGCCCATCTCCC.............CAAGGC.......... 215

     58 hoTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAla 74
                        |||   |||
    216 ..........CGACATGACTGCCTC...................... 230

     75 AlaLysAlaProGlyAlaSer...SerArgGluGluProProGluGluPr 90
                  |||   |||      |||      |||   :::::||
    231 ...........GGGGACTCCTCACTACGGGCAGCCCCACAAGATCAACC 268

     90 aGluGlnCysProVal...IleAspSerGlnAlaProAlaGlySer.... 104
        |            :::   :::|:::::::|||      ||:::
    269 CCCTCCCACCACAGCAGGACTGGATCAATCAGCCAGTCGAGGCTCAACGTC 318

    105 .........................LeuAspLeuValProGlyGlyLeu 112
                                 ::::::   |||   |||||||||
    319 AAGCGGGAGTATGACCACATGAATGGATCCAGGGAGTCTCCGGTGGACTG 368

    113 .....................................ThrLeuGluGluHi 117
                                           ::::::   ||
    369 CAGCGTTAGCAAATGCAGCAAGCTGGTGGGCGGAGGCGAGTCCAACCCCA 418

    117 sSerLeuGluGlnValGlnSerMetValVal.GlyGluValLeuLysAsp 133
        |:::|||:::|||:::   :::         |||            :::
    419 TGAACTACAACAGCTATATGGACGAGAAGAATGGCCCCCCTCCTCCCAAC 468

    134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTr 150
        :::   |||     :::   :::   :::|||||||      ||
    469 ATGACCACCAACGAGAGGAGAGTCATCGTCCCCGCAGACCCCACACTGTG 518

    150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrA 167
        |:::   :::::|||:::::::|||||   ||:::      :::|||
    519 GACACAGGAGCATGTGAGGCAATGGCTGGAGTGGGCCATAAAGGAGTATA 568

    167 rqLeuProProMetGlyLysAla...PheGluGluLeuAlaGlyLysGlu 182
          |||       :::       :::   ||||||:::::   |||||||||
    569 GCTTGATGGCAGATCGACACATCCTTTTCCAGAACATGGATGGCAAGGAA 618

    183 LeuCysAlaMetSerGluGluGlnPheArgGluArgSerProLeu...Gl 198
        ||||||   ||:::::::|||:::|||   :::   :::   |||   ::
    619 CTGTGTAAAATGAACAAGGAGGACTTCCTCCGCGCCACCACCCTCTACAA 668

    198 yGlyAspValLeuHisAlaHisLeu..................... 206
        :   ::::|||||||   :::|||||
    669 CACGCAACTGCTGTTGTCACACCTCAGTTACCTCAGGCAAAGTTCACTGC 718

    206 ..................................................... 206

    719 TGGCCTATAATACAACCTCCCACACCGACCAATCCTCACGATTGAGTGTC 768

    207 .................AspIleTrpLysSerAlaAlaTrpMetLysGl 217
                        |||      :::   :::|||||||   :::::
```

```
     769 AAAGAAGACCCTTCTTATGACTCAGTCAGAAGAGGACCATGGGGCAATAA 818

     217 uArgThrSerPro.....GlyAlaIleHisTyrC 227
         :   :::|||                 |||:::   :::
     819 CATGAATTCTGGCCTCAACAAAAGTCCTCCCCTTGGAGGGGCACAAACGA 868

     227 ysAlaSerThrSerGluGlu..... 233
           ::::::::::::|||:::
     869 TCAGTAAGAATACAGAGCAACGGCCCCAGCCAGATCCGTATCAGATCCTG 918

     234 SerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHi 250
         :::   |||:::|||::::::   :::      |||||||||   |||::
     919 GGCCCGACCAGCAGTCGCCTAGCCAACCCTGGAAGCGGGCAG...ATCCA 965

     250 sLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyA 267
         :|||||||||||||||   |||   |||||||::      :::   :::
     966 GCTGTGGCAATTCCTCCTGGAG...CTGCTCTCCGACAGCGCCAACGCCA 1012

     267 rqPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAsp 283
           |||   |||   :::      :::|||   ||||||:::   |||
    1013 GCTGTATCACCTGG...GAGGGACCAACGGGGAGTTCAAAATGACGGAC 1059

     284 SerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMe 300
           :::||||||||||   ||||||   |||||:::::::|||   ||
    1060 CCCGATGAGGTGGCCAGGCGCTGGGGCGAGCGGAAAAGCAAGCCCAACAT 1109

     400 tAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyI 417
         |||||||||||||||||||||||:::::|||   ||||||   |||:::|
    1110 GAATTACGACAAGCTGAGCCGGGCCCTCCGTTATTACTATGATAAAAACA 1159

     417 leIleArgLysProAspIleSerGlnArgLeuValTyrGlnPhe 431
         ||:::   |||          :::::|||   :::|||:::|||
    1160 TTATGACCAAAGTGCAC...GGCAAAAGATATGCTTACAAATTT 1200

seq_name: qb_pat1:AR080101

seq_documentation_block:
LOCUS       AR080101     2938 bp    DNA              PAT       31-AUG-2000
DEFINITION  Sequence 3 from patent US 5968734.
ACCESSION   AR080101
VERSION     AR080101.1  GI:10006836
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2938)
  AUTHORS   Aurias,A., Delattre,O., Desmaze,C., Melot,T., Peter,M.,
            Plouquastel,B., Thomas,G. and Zucman,J.
  TITLE     Nucleic acid corresponding to a gene of chromosome 22 involved in
            recurrent chromosomal translocations associated with the
            development of cancerous tumors, and nucleic acids of fusion
            resulting from said translocations
  JOURNAL   Patent: US 5968734-A 3 19-OCT 1999;
FEATURES             Location/Qualifiers
     source          1..2938
                     /organism="unknown"
BASE COUNT      847 a    692 c    675 g    724 t
ORIGIN


alignment scores:
            Quality:  279.50          Length:     382
              Ratio:  1.522             Gaps:      18
 Percent Similarity:  47.906   Percent Identity: 28.010

alignment block:
 US-09-126-945B-2 x AR080101  ..

 Align seq 1/1  to: AR080101  from: 1  to: 2938

      34 GlyAlaValGlyLeuGluArg.....Ar 41
         |||:::|||||   ||||||          ||
     163 GGCTCTCTCCGGTGGTGAGCGACCACGTCCCTCTTTGACTCAGCGTACG 212

  41 qAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerAlaP 58
     |:::   ||||||||||||              ||||||
 213 GAGCGGCAGCCCATCTCCC.............CAAGGC..... 237

  58 heTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAla 74
                    |||   |||
 238 ...........CGACATGACTGCCTC..... 252

  75 AlaLysAlaProGlyAlaSer...SerArgGlnGluProProGluGluPr 90
             |||   |||      |||      |||   :::::::||
 253 ...........GGGGAGTCCTGACTACGGGCAGCCCCACAAGATCAACC 290

  90 oGluGlnCysProVal...IleAspSerGlnAlaProAlaGlySer.... 104
     |            :::   :::|||::::::|||      |||:::
 291 CCCTCCCACCACAGCAGGAGTGGATCAATCAGCCAGTGAGGGTCAACGTC 340

 105 .........................LeuAspLeuValProGlyGlyLeu 112
                                  ::::::   |||   |||||||||
 341 AAGCGGGACTATGACCACATGAATGGATCCAGGGAGTCTCCGGTGGACTG 390

 113 .....................................ThrLeuGluGluHi 117
                                          ::::::   ||
 391 CAGCGTTAGCAAATGCAGCAAGCTGGTGGGCGGAGGCGAGTCCAACCCCA 440

 117 sSerLeuGluGlnValGlnSerMetValVal.GlyGluValLeuLysAsp 133
     |:::|||:::|||:::   :::          |||            :::
 441 TGAACTACAACAGCTATATGGACGAGAAGAATGGCCCCCCTCCTCCCAAC 490

 134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTr 150
     :::   |||     :::   :::   :::  |||||||||       ||
 491 ATGACCACCAACGAGAGGAGAGTCATCGTCCCCGCAGACCCCACACTGTG 540

 150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrA 167
     |:::   :::::|||:::::|||||   |||'''       ''::|||
 541 GACACAGGAGCATGTGAGGCAATGGCTGGAGTGGGCCATAAAGGAGTATA 590

 167 rqLeuProProMetGlyLysAla...PheGlnGluLeuAlaGlyLysGlu 182
       |||      :::      :::   ||||||::::::   |||||||||
 591 GCTTGATGGAGATCGACACATCCTTTTTCCAGAACATGGATGGCAAGGAA 640

 183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeu...Gl 198
     ||||||   |||:::::|||:::|||   :::   :::   |||   ::
 641 CTGTGTAAAATGAACAAGGAGGACTTCCTCCGCGCCACCACCCTCTACAA 690

 198 yGlyAspValLeuHisAlaHisLeu..... 206
     :  :::||||||   :::||||||
 691 CACGGAAGTGCTGTTGTCACACCTCAGTTACCTCAGGGAAAGTTCACTGC 740

 206 ..... 206

 741 TGGCCTATAATACAACCTCCCACACCGACCAATCCTCACGATTGAGTGTC 790

 207 ..          ....AspIleTrpLysSerAlaAlaTrpMetLysGl 217
                    |||      :::   :::||||||   :::::
 791 AAAGAAGACCCTTCTTATGACTCAGTCAGAAGAGGAGCATGGGGCAATAA 840

 217 uArgThrSerPro.....GlyAlaIleHisTyrC 227
     :   :::|||                 |||:::   :::
 841 CATGAATTCTGGCCTCAACAAAAGTCCTCCCCTTGGAGGGGCACAAACGA 890

 227 ysAlaSerThrSerGluGlu..... 233
       :::::::::::::|||:::
 891 TCAGTAAGAATACAGAGCAACGGCCCCAGCCAGATCCGTATCAGATCCTG 940

 234 SerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHi 250
     :::   |||:::|||::::::   :::      |||||||||   |||::
 941 GGCCCGACCAGCAGTCGCCTAGCCAACCCTGGAAGCGGGCAG...ATCCA 987

 250 sLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyA 267
     :|||||||||||||||   |||   |||||||::      :::   :::
 988 GCTGTGGCAATTCCTCCTGGAG...CTGCTCTCCGACAGCGCCAACGCCA 1034
```

```
     267 rqPheIleArqIrpLeuAsnLysGluLysGlyIlePheLysIleGluAsp 284
         ||:  |||  :::     :::|||  ||||||:::  |||
    1035 GCTGTATTACCTGG...GAGGGGACCAACGGGGAGTTCAAAATGACGGAC 1081

     284 SerAlaGlnValAlaArqLeuIrpGlyIleArqLysAsnArqProAlaMe 300
         :::|||||||||  ||||||  |||||||:::::|||  ||
    1082 CCCGATGAGGTGGCCAGCGGCTGGGGCGACGGGAAAAGCAACGCCCAACAT 1131

     300 tAsnTyrAspLysLeuSerArgSerIleArqGlnTyrTyrLysLysGlyI 317
         ||||||||||||||||||||||:::::|||  ||||||  |||:::|
    1132 GAATTACCACAAGCTCAGCGGGGCCCTCCGTTATTACTATGATAAAAACA 1181

     317 leIleArqLysProAspIleSerGlnArqLeuValTyrGlnPhe 331
         ||:::  |||       :::::|||  :::|||:::|||
    1182 ITATGACCAAAGTGCAC...GGCAAAAGATATGCTTACAAATTT 1222

seq_name: qb_pr9:HSHUMFLI

seq_documentation_block:
LOCUS       HSHUMFLI     2938 bp    mRNA            PRI       28 JUN 1995
DEFINITION  H.sapiens HUMFLI-1 mRNA.
ACCESSION   X67001 S44250
VERSION     X67001.1  GI:42529
KEYWORDS    FLI-1 gene homologue.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2938)
  AUTHORS   Delattre,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1992) O. Delattre, Lab. de Genet. des Tumeurs
            Inst. Curie, 26 rue D'Ulm, 75231 Paris Cedex, FRANCE
REFERENCE   2  (bases 1 to 2938)
  AUTHORS   Delattre,O., Zucman,J., Plougastel,B., Desmaze,C., Melot,T.,
            Peter,M., Kovar,H., Joubert,I., de Jong,P., Rouleau,G., Aurias,A.
            and Thomas,G.
  TITLE     Gene fusion with an ETS DNA-binding domain caused by chromosome
            translocation in human tumours
  JOURNAL   Nature 359 (6391), 162-165 (1992)
  MEDLINE   92396239
FEATURES             location/Qualifiers
     source          1. .2938
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11q24"
                     /dev_stage "adult"
                     /tissue_type="bone marrow"
                     /clone_lib="cDNA, Clontech HL1058"
     gene            143. .1501
                     /gene "HUMFLI-1"
     CDS             143. .1501
                     /gene="HUMFLI-1"
                     /codon_start 1
                     /product="homologue of the murine  FLI 1 gene"
                     /protein_id="CAA47399.1"
                     /db_xref="GI:42530"
                     /db_xref "SWISS-PROT:q01543"
                     /translation="MEGTIKEAISVVSEDQSLFDSAYGAAAHIPKADMTASGSPDYGQ
                     PHKINPLPPVQQFEWTNQFVSVNVKELYDHMNGSSLSFYIPKVSKSKLVGGLDSNFMNY
                     NSYMDEKNGPPPPNMITTNERRVIVPADPTLWTQEHVEQWLEWAIKEYSLMEIDTSFFQ
                     NMDGKELCKMNKEDFLRATTLYNTFVILSHLSYLRESSILAYNITSHTIQSSRLSVKF
                     DPSYDSVRRGAWGNNMNSGLNKSPPLGGAQTISKNTEQRPQPDPYQTLGPTSSRLANP
                     GSGQTQLWQFLLELLSDSANASCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDK
                     LSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPPESSMYKYPSDISYMPSYH
                     AHQQKVNFVPPHPSSMPVTSSSFFGAASQYWTSPTGGIYPNPNVPRHPNTHVPSHLGS
                     YY"
     misc_feature    971. .1222
                     /gene="HUMFLI-1"
                     /note="ETS domain"
     polyA_signal    2890. .2895
     polyA_signal    2411. .2416
     polyA_signal    2415. .2420
     polyA_signal    2908. .2913
                     /evidence experimental
     polyA_site      2927
BASE COUNT      847 a    692 c    675 g    724 t
ORIGIN


alignment_scores:
           Quality:  278.50              Length:     382
             Ratio:  1.522                 Gaps:      18
 Percent Similarity:  47.906   Percent Identity: 28.010

alignment_block:
 US-09-126-945B-2 x HSHUMFLI  ..

 Align seq 1/1  to: HSHUMFLI  from: 1  to: 2938

      34 GlyAlaValGlyLeuGluArq.....................Ar 41
         |||:::|||||  :|||||                      ||
     163 GGCTCTGTCGGTGGTGAGCGACGACCAGTCCCTCTTTGACTCAGCGTACG 212

      41 qAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerAlaP 58
         |:::  |:||||:|||||                 ||||||
     213 GAGCGGCAGCCCATCTCCC.............CAAGGC......... 237

      58 heTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAla 74
                       |||  |||
     238 ..........CGACATGACTGCCTC..................... 252

      75 AlaLysAlaProGlyAlaSer...SerArqGluGluProProGluGluPr 90
                 |||  |||  |||     |||  :::::::||
     253         GGGGAGTCCTGACTACGGGCAGCCCCACAAGATCAACC 290

      90 oGluGlnCysProVal...IleAspSerGlnAlaProAlaGlySer.... 104
         |        :::  :::|||||||||:|||     |||:::
     291 CCCTCCCACCACAGCAGGAGTGGATCAATCAGCCAGTGAGGCTCAACGTC 340

     105 .........................LeuAspLeuValProGlyGlyLeu 112
                                  ::::::  |||  |||||||||
     341 AAGCGGCAGTATGACCACATGAATGGATCCAGGGAGTCTCGGGTGGACTG 390

     113 .................................ThrLeuGluGluHi 117
                                         ::::::  ||
     391 CAGCGTTAGCAAATGCAGCAAGCTGGTGGGCGGAGGCGAGTCCAACCCCA 440

     117 sSerLeuGluGlnValGlnSerMetValVal.GlyGluValLeuLysAsp 133
         |:::|||:::|||:::  :::      |||          :::
     441 TGAACTACAACAGCTATATGGACGAGAAGAATGGCCCCCCTCCTCCCAAC 490

     134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIr 150
         :::  |||     :::  :::  :::  |||||||||     ||
     491 ATGACCACCAACGACAGGAGAGTCATCGTCCCCGGCAGCACCCACACTGTG 540

     150 pSerProSerAsnValGluLysIrpLeuLeuIrpThrGluHisGlnTyrA 167
         |:::  :::::|||:::::|||||  |||:::      :::|||
     541 GACACAGGAGCATGTGAGGCAATGGCTGGAGTGGGCCATAAAGGATATA 590

     167 rqLeuProProMetGlyLysAla...PheGlnGluLeuAlaGlyLysGlu 182
           |||      :::      :::   |||||:::::   |||||||||
     591 GCTTGATGGAGATGACACATCCTTTTCCAGAACATGGATGGCAAGGAA 640

     183 IouCysAlaMetSerGlnGluGlnPheArqGlnArqSerProLeu...Gl 198
         ||||||  |||:::::::|||:::|||  :::  :::  |||  ::
     641 CTGTGTAAAATAACAAGAGTACTCCTCCGCGACACACCTCTACAA 690

     198 yGlyAspValLeuHisAlaHisLeu..................... 206
         :  :::|||||  :::|||||
     691 CACGGAAGTGCTGTTGTCACACCTCAGTTACCTCAGGGAAAGTTCACTGC 740

     206 ..................................................... 206

     741 GGCCTATAATACAACCTCTCACACTGACCAATCCTCACGATGAGTGTC 790
```

```
     207 ..................AspIleTrpLysSerAlaALaTrpMetLysGl 217
                           |||       :::   :::||||||   :::::
     791 AAAGAAGACCCTTCTTATGACTCAGTCAGAAGAGGAGCATGGGGCAATAA 840

     217 qArgThrSerPro.....................GlyAlaIleHisTyrC 227
         :   :::|||                       |||:::   :::
     841 CATGAATTCTGGCCTCAACAAAAGTCCTCCCCTTGGAGGGGCACAAACGA 890

     227 ysAlaSerThrSerGlaGlu.............................. 241
         :::::::::::|||:::
     891 TCAGTAAGAATACAGAGCAACGGCCCCAGCCAGATCCGTATCAGATCCTG 940

     241 SerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHi 260
         :::   |||:::|||:::::   :::       ||||||||   ||:::
     941 GGCCCCACCAGCAGTCGCCTAGCCAACCCTGGAAGCGGGCAG...AICCA 987

     261 sLeuTrpGlaPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyA 277
         :|||||||||||||||   |||   ||||||:::      :::  :::
     988 GCTGTGGCAATTCCTCCTGGAG...CTGCTCTCCGACAGCGCCAACGCCA 1034

     277 rqPheIleArqTrpLeuAsnLysGluLysGlyIlePheLysIleGluAsp 293
         |||   |||  :::      :::|||   ||||||:::   |||
    1035 GCTGTATCACCTGG...GAGGGGACCAACGGGGAGTTCAAAATGACGGAC 1081

     294 SerAlaGlnValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMe 310
         :::||||||||   ||||||   ||||||:::::::||   ||
    1082 CCCGATGAGGTGGCCAGGCGCTGGGGCGAGCGGAAAAGCAAGCCCAACAI 1131

     311 tAsnTyrAspLysLeuSerArqSerIleArqTyrTyrLysLys:GlyI 327
         ||||||||||||||||||||||:::::|||   ||||||   |||:::|
    1132 CAATTACGACAAGCTGAGCCGGGCCCTCCGTTATTACTATGATAAAAAC'A 1181

     327 IeIleArqLysProAspIleSerGlnArqLeuValTyrGlnPhe 341
         ||:::   |||         :::::|||   :::|||:::||
    1182 TTATGACCAAAGTGCAC...GGCAAAAGATATGCTTACAAATTT 1222

Seq name: gb pri0:HUMERGBFLI

Seq documentation block:
LOCUS       HUMERGBFLI    2957 bp    mRNA            PRI       24 FEB 2000
DEFINITION  Homo sapiens ERGB transcription factor mRNA, complete cds.
ACCESSION   M98833
VERSION     M98833.4  GI:7025922
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2957)
  AUTHORS   Watson,D.K., Smyth,F.E., Thompson,D.M., Cheng,J.Q., Testa,J.R.,
            Papas,T.S. and Seth,A.
  TITLE     The ERGB/FLI1 gene: Isolation and characterization of a new member
            of the family of human ETS transcription factors
  JOURNAL   Cell Growth Differ. 3, 705-713 (1992)
  MEDLINE   93075640
REFERENCE   2  (bases 1 to 2957)
  AUTHORS   Watson,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18 AUG 1992) Center for Molecular and Structural
            Biology, Medical University of South Carolina, 86 Jonathan Lucas St
            PO Box 250956, Charleston, SC 29425, USA
REFERENCE   3  (bases 1 to 2957)
  AUTHORS   Watson,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24 FEB 2000) Center for Molecular and Structural
            Biology, Medical University of South Carolina, 86 Jonathan Lucas St
            PO Box 250956, Charleston, SC 29425, USA
  REMARK    Sequence update by submitter
COMMENT     On Feb 24, 2000 this sequence version replaced gi:7025920.
FEATURES             Location/Qualifiers
     source          1. .2957
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /map "Chromosome 11, awaiting qualifier approval"
                     /cell_line="CEM"
                     /cell_type "T-cell"
                     /tissue_lib="lambda gt10"
     CDS             173. .1541
                     /note "similar to FLI-1"
                     /codon_start=1
                     /product "ERGB transcription factor"
                     /protein_id "AAA35812.2"
                     /db_xref "GI:7025921"
                     /translation "MDGIIKEALSVVSDDQSLFDSAYGAAAHLPKADMTASGSPDYGQ
                     PHKINPLPPQQEWINQPVPVNVKREEYDRMNGSRESPVDCSVSKCSKLVGGGESNPMNY
                     NSYMDEKNCPPPPNMTTNERRVIVPADPTLWSTQHVPQWIEWATKEYSLMEIDTSFFQ
                     NMDGKELCKMNKEDFLRALTLYNTEVLLSHLSYLREASLLAYNTISHTDQSSKLSVKE
                     DPSYDSVRRGAWGNNMNSGLNKSPPLGGAQLTSKNTEQKPQPDPYQILGPTSSRLANP
                     GSGQIQLWQFLLELLSDSANASCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDK
                     LSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPTESSMYKYPSDISYMPSYH
                     AHQQKVNFVPPHPSSMPVTSSSFFGAASQYWTSPTGGIYPNPNVPRHPNTHVPSHLGS
                     YY"
     polyA_signal    2925. .2930
BASE COUNT      855 a    698 c    680 g    724 t
ORIGIN


alignment_scores:
            Quality:  278.50              Length:     482
              Ratio:  1.522                 Gaps:      18
 Percent Similarity:  47.906   Percent Identity: 28.010

alignment block:
 US-09-126-945B-2 x HUMERGBFLI   ..

 Align seq 1/1  to: HUMERGBFLI  from: 1  to: 2957

      34 GlyAlaValGlyLeuGluArq.........................Ar 41
         |||::|||||   ||||||                           ||
     193 GGCTCTGTCGGTGGTGAGCGACGACCAGTCCCTCTTTGACTCAGCGTACG 242

      41 qAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerAlaP 58
         |:::   ||||||||||||              ||||||
     243 GAGCGGCAGCCCATCTCCC.............CAAGGC.......... 267

      58 heTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAla 74
                        ||   |||
     268 ..........CGACATGACTGCCTC....................... 282

      75 AlaLysAlaProGlyAlaSer...SerArqGlnGluProProGlnGluPr 90
                 |||   |||     |||     |||   :::::::||
     283 ...........GGGGAGTCCTGACTACGGGCAGCCCCACAAGATCAACCC 320

      90 qGluGlnCysProVal...IleAspSerGlnAlaProAlaGlySer.... 104
         |          :::   :::|||:::::::|||     |||:::
     321 CCCTCCCACCACAGCAGGAGTGGATCAATCAGCCAGTGAGGGTCAACCGTC 370

     105 .........................LeuAspLeuValProGlyGlyLeu 112
                                  ::::::   |||   |||||||||
     371 AAGCGGGAGTATGACCACATGAATGGATCCAGGGAGTCTCCGGTGGACTG 420

     113 .....................................ThrLeuGluGluHi 117
                                          ::::::  ||
     421 CAGCGTTAGCAAATGCAGCAAGCTGGTGGGCGGAGGCGACTCCAACCCCA 470

     117 sSerLeuGluGlnValGlnSerMetValVal.GlyGluValLeuLysAsp 133
         |:::|||:::|||:::   :::         |||                :::
     471 TGAACTACAACAGCTATATGGACGAGAAGAATGGCCCCCCTCCTCCCAAC 520

     134 IleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspIr 150
         :::   |||     :::   :::   :::   |||||||||      ||
     521 ATGACCACCAACGAGAGGAGAGTCATCGTCCCCGCAGACCCCACACTGTG 570

     150 pSerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrA 167
         |:::   :::::::|||:::::::|||||   |||:::      :::|||
     571 GACACAGGAGCATGTGAGGCAATGGCTGGAGTGGGCCATAAAGGAGTACA 620
```

```
     167 rqLeuProProMetGlyLysAla...PheGlnGlnLeuAlaGlyLysGlu 182
           |||       :::       :::   ||||||||||:::   |||||||||
     621 GCTTGATGGAGATCGACACATCCTTTTCCAGAACATGGATGGCAAGGAA 670

     183 LeuCysAlaMetSerGluGluGlnPheArgGlnArgSerProLeu...Gl 198
         ||||||   |||::::::|||:::|||   :::   :::   |||   ::
     671 CTGTGTAAAATGAACAAGGAGGACTTCCTCCGCGCCACCACCCTCTACAA 720

     198 yGlyAspValLeuHisAlaHisLeu..........................  207
         :   :::||||||   :::|||||
     721 CACGGAAGTGCTGTTGTCACACCTCACTTACCTCAGGCAAAGTTCACTGC 770

     206 ..................................................  206

     771 TGGCCTATAATACAACCTCCCACACCGACCAATCCTCACGATTGACTGTC 820

     207 ..................AspIleTrpLysSerAlaAlaTrpMetLysGl 217
                            |||       :::   :::||||||   :::::
     821 AAAGAAGACCCTTCTTATGACTCACTCAGAAGAGGAGCTTGGGGCAATAA 870

     217 qArqThrSerPro.....................GlyAlaIleHistyr' 227
         :   :::|||                       ||:::   :::
     871 CATGAATTCTGGCCTCAACAAAAGTCCTCCCCTTGGAGGGGCACAAACGA 920

     227 ysAlaSerThrSerGluGlu.............................. 242
           :::::::::::::|||:::
     921 TCAGTAAGAATACAGACCAACGGCCCCAGCCAGATCCGTATCAGATCCTG 970

     243 SerTrpThrAspSerGluValAspSerSerCysSerGlyGlnProIleHi 262
         :::   | |:::|||::::::  :::     ||||||||||   |||::
     971 GGCCCGACCAGCAGTCGCCTAGCCAACCCTGGAAGCGGGCAG...ATCCA 1017

     263 sLeuTrpGlnPheLeuLysGlnLeuIleGluLysPr:HisSerTyrGlyA 282
         :|||||||||||||||   |||   ||||||:::     :::   :::
    1018 GCTGTGGCAATTCCTCCTGGAG...CTGCTCTCCGACAGCGCCAACCGCA 1064

     283 rqPheIleArqTrpLeuAsnLysGluLysGlyIleMetLysIleGlnAsp 302
           ||.   |||   :::       :::|||   ||||||:::   |||
    1065 GCTGTATCACCTGG...GAGGGGACCAACGGGGAGTTCAAAATGACGGAC 1111

     303 SerAlaGlnValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMe 322
           :::||||||||   ||||||   ||||||:::::::||   ||
    1112 CCCGATGAGGTGGCCAGCGCCTGGGGCGAGCGGAAAAGCAAGCCTAACAT 1161

     323 tAsnTyrAspLysLeuSerArqSerIleArqGlnTyrTyrLysLysGlyI 342
         |||||||||||||||||||||||:::::|||   ||||||   ||:::|
    1162 GAATTACGACAAGCTGAGCCGGCGCCTCCGTTATTACTATGATAAAAGCA 1211

     343 leIleArqLysProAspIleSerGlnArqLeuValTyrGlnPhe 441
         |||::   |||         :::::|||   :::|||:::|||
    1212 TTATGACCAAAGTGCAC...GGCAAAAGATATGCTTACAAATTT 1252

seq_name: qb_pr7:BC001670

seq_documentation_block:
LOCUS       BC001670     2959 bp    mRNA            PRI       16-MAR-2001
DEFINITION  Homo sapiens, Friend leukemia virus integration 1, clone MGC:2252,
            mRNA, complete cds.
ACCESSION   BC001670
VERSION     BC001670.1  GI:12804518
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2959)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16 JAN 2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            Contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Kettleman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at:  http://image.llnl.gov
            Series: IRAL Plate: 3 Row: i Column: 14.
FEATURES             Location/Qualifiers
     source          1. .2959
                     /organism "Homo sapiens"
                     /db_xref "LocusID:2314"
                     /db_xref "Taxon:9606"
                     /clone "MGC:2252 IMAGE:2989502"
                     /tissue_type "Lymph, Burkitt lymphoma"
                     /clone_lib "NIH_MGC_8"
                     /lab_host "DH10B-R"
                     /note="Vector: pOTB7"
     CDS             124. .1482
                     /codon_start 1
                     /product="Friend leukemia virus integration 1"
                     /protein_id "AAH01670.1"
                     /db_xref "GI:12804519"
                     /translation "MEGLIKEALSVVSDQQSLEDSAYGAAAHLPAPMTASGSLLYGL
                     PHKLNPLEPQQEWLNGEVSVNVEKRLPHMNGSGLDSRVPV SVSSGSKLVGSGENFMNY
                     NSYMDGENGFLFLNMLLNEEKV.VEALFLGWL.PDVEQWLLGWAIEEYGLMPLLISFFG
                     NMDGKELGKMNKIDPLRATFLYNIFVLLSHLSYLRFSSLLAYNITSHLQSSKLSVKE
                     DPSYDSVRPGAWGNNMNSGLNKSPPLGGAQLTSENIEQRPQPDPYQTLGPISSRLANP
                     GSGQTQLWQFLLELLSDSANASCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDK
                     LSRALRYYYDRNIMTKVHGKRYAYRFDFHGIAQALQPHLTTSFGKYEVTSDISYMPSYH
                     AHQQKVNFVADNDSSMPVTSSSFPGAASQVWVSPTGGIYPNPNVPRRFNIRVPSHLSS
                     YY"
BASE COUNT       885 a    689 c    663 g    722 t
ORIGIN


alignment_scores:
            Quality:  278.50              Length:    382
              Ratio:  1.522                Gaps:      18
 Percent Similarity:  47.906   Percent Identity: 28.010

alignment_block:
 US-09-126-945B-2 x BC001670   ..

 Align seq 1/1  to: BC001670  from: 1  to: 2959

      34 GlyAlaValGlyLeuGlnArq.............................Ar 41
         ||||::|||||   ||||||                               ||
     144 GGCTCTGTCGGTGGTGAGCGACGACCAGTCCCTCTTGACTCAGCGTACG 193

      41 qAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerAlaP 58
         |:::   ||||||||||||                 ||||||
     194 GAGCGGCAGCCCATCTCCC..............CAAGGC.......... 218

      58 heTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAla 74
                         |||   |||
     219 ..........CGACATGACTGCCTC.......................... 233

      75 AlaLysAlaProGlyAlaSer...SerArqGlnGluProProGluGluPr 90
                    |||   |||       |||       |||   :::::||
     234 ..........GGGGAGTCCTGACTACGGGCAGCCCCACAAGATCAACC 271

      90 oGluGlnCysProVal...IleAspSerGlnAlaProAlaGlySer.... 104
         |       :::   :::|||::::::|||       ||:::
     272 CCCTCCCACCACAGCAGGAGTGGATCAATCAGCCAGTGAGGGTCAACGTC 321
```



```
seq_name: qb rol:MUSGAA

seq_documentation_block:
LOCUS       MUSGAA          1965 bp    mRNA             rod       09 NOV 1994
DEFINITION  Mouse GA binding protein (GABP alpha subunit) mRNA, complete cds.
ACCESSION   M74515 M74511
VERSION     M74515.1  GI:193382
KEYWORDS    GA binding protein.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1965)
  AUTHORS   LaMarco,K., Thompson,C.C., Byers,B.P., Walton,E.M. and
            McKnight,S.L.
  TITLE     Identification of Ets- and notch-related subunits in GA binding
            protein [see comments]
  JOURNAL   Science 253 (5021), 789-792 (1991)
  MEDLINE   91343912
FEATURES             Location/Qualifiers
     source          1..1965
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="3T3-L1"
     gene            454..1818
                     /gene="GABP"
     CDS             454..1818
                     /gene="GABP"
                     /note="alpha subunit"
                     /codon_start=1
                     /product="GA binding protein"
                     /protein_id="AAA53040.1"
                     /db_xref="GI:193383"
                     /translation="MTKREAEELIETEIDGTEKAECTEESIVEQTYTPAECVSQAIDI
                     NEPIGNLKKLLEPKIQCSLDABEICLQDIQLLDPURSLEIQQVKIIGTVQLSVQVISYQ
                     GMEEKLNILETVKTAETVEVVIDPDAHHAEAEAHLVEEAQVIILDGTKHTIISDETS
                     EQVTRWAAALEGYRKEQERLGIPYDPIRWSTDQVLHWVVWVMKEFSMTDIDLITLNIS
                     GRELCSLNQFDEFQRVPRGFILWSHLFLIRKYVLASQEQQMNEIVTIDQPVQIIPASV
                     PPATPITIKVINSSAKAAKVQRSPRISGEDRSSPGNRTGNNGQIQLWQFLLELLIDKD
                     ARDCISWVGHEGEFKLNQPELVAQKWCQRENKPTMNYEKLSRALRYYYIGDMIVKVQG
                     FRFVYRFVCDNKTLIGYSAAFLNRIVIPCFQKPLARMQLDGIAQPVIAVALAAISLQA
                     LKEL"
BASE COUNT      545 a    431 c    549 g    450 t
ORIGIN


alignment_scores:
            Quality:  277.00              Length:    288
              Ratio:  1.753                 Gaps:      9
 Percent Similarity:  54.861   Percent Identity: 28.819

alignment_block:
 US-09-126-945B-2 x MUSGAA    ..

 Align seq 1/1  to: MUSGAA  from: 1  to: 1965

      78 ProGlyAlaSerSerArgGluGluProProGluGluProGluGlnCysPr  94
         |||   |||       |||       :::       ||||::::::
     829 CCAGATGCCCACCACGCGGAAGCAGAAGCGCATCTCGTTGAAGAAGCTCA 878

      94 oValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyG 111
         ||||||                    ::||||||   :::
     879 AGTGATA....................ACTCTTGACGGCACCAAGCACA 907

     111 lyLeuThrLeuGluGluHisSerLeuGluGlnValGluSerMetValVal 127
         ||||||||||||||||||   ||||||||         ::::::
     908 TTACGACCATTCAGACGAGACTCTGGAGCAGGTGACGACATGGGCTGCT 957

     128 GlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleTh 144
         ::||||::   :::       :::    ||||::|||
     958 .....GCACTGGAAGGCTACAGAAAAGAGCAGGAGCGCCTGGCATCCC 1001

     144 rAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuT 161
```

```
          ||||||:::   ||||||   ::::::|||      ||::::::|
1002 CTATGATCCTATACGCTGGTCCACGGACCAAGTCCTGCATTGGGTGGTTT 1051

 161 rpThrGluHisGluTyrArgLeuProProMetGlyLysAlaPheGluGlu 177
     ||:::     ::::::  :::       :::      :::      :::
1052 GGGTAATGAAGCACTTCAGCATCACTGATATAGACCTCACCACACTCAAC 1101

 178 LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnAl 194
     ::::::|||:::||||||||::::::::::::::|||:::|||   |||||
1102 ATTTCGGGAAGAGAATTATGTAGTCTCAACCAAGAAGATTTTTTTCAGCG 1151

 194 gSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysS 211
     |   |||        |||:::::|||    :::|||||:::::   ::::
1152 GGTCCCT...CGGGGAGAAATTCTTTGGAGTCATCTGGAGCTTCTTCGAA 1198

 211 erAlaAlaTrpMetLysGluArg........................... 218
     ::    :::      :::::::::
1199 AATATGTTTTGGCCAGCCAAGAGCAACAGATGAATGAGATAGTTACCATT 1248

 219 ......................................ThrSerProGl 222
                                            :::::::|||
1249 GACCAGCCTGTGCAGATTATTCCAGCCTCAGTGCCTCCCGCTACACCGAC 1298

 222 yAlaIleHisTyrCysAlaSerThrSer..............GluGlnG 234
       :::|||          |||:::::                  :::::::
1299 TACGATTAAAGTTATAAACAGCAGTGCAAAAGCAGCTAAAGTGCAACGGT 1348

 234 erTrpThrAspSerGluValAspSerSerCysSerGlyGln......... 247
      ||          |||        |||   |||         |||:::
1349 CCCCAAGGATTTCAGGAGAAGACAGAAGTTCACCGGGGAACAGAACAGGA 1398

 248 .........ProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLy 261
                 |||:::|||||||||||||||     |||||||||
1399 AACAATGGTCAGATCCAACTATGGCAGTTTTTGCTAGAACTTCTTACTGA 1448

 261 sProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyI 278
          :::               |||   |||::::::   |||    |||
1449 CAAGGATGCTCGAGAC...TGTATTTCTTGGGTTGGTGATGAA...GGTG 1492

 278 lePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArg 294
         ||||||:::::::::        |||||||:::      ||||||    |||
1493 AATTTAAGCTAAATCAGCCTGAGTTGGTTGCGCAAAAATGGGGACAACGT 1542

 295 LysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGl 311
     ||||||:::|||:::||||||||:::|||||||||||||||:::::::|||
1543 AAGAACAAGCCTACCATGAACTATGAGAAACTTAGCCGTGCATTACGGTA 1592

 311 nTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuV 328
       |||||          :::|||   |||    :::    ::::::|||:::|
1593 TTATTATGATGGGGACATGATTTGTAAAGTTCAA...GGCAAGAGATTTG 1639

 328 alTyrGlnPheVal 332
      |||||::::|||||
1640 TGTACAAATTTGTT 1653
```

seq_name: gb_rol:AF256217

```
seq_documentation_block:
LOCUS       AF256217     2688 bp    mRNA            ROD       02-MAY-2000
DEFINITION  Mus musculus ets family transcription factor ELF2A2 (ELF2) mRNA,
            complete cds, alternatively spliced.
ACCESSION   AF256217
VERSION     AF256217.1  GI:7677533
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2688)
  AUTHORS   Wilkinson,D.A., Neale,G.A., Mao,S., Naeve,C.W. and Goorha,R.M.
  TITLE     Elf-2, a rhombotin-2 binding ets transcription factor: discovery
            and potential role in T cell leukemia
  JOURNAL   Leukemia 11 (1), 86-96 (1997)
  MEDLINE   97154704
REFERENCE   2  (bases 1 to 2688)
  AUTHORS   Wilkinson,D.A., Neale,G.A.M., Mao,S., Fernandes,E.R.,
            Davenport,J.W., Naeve,C.W. and Goorha,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-2000) Virology and Molecular Biology, St. Jude
            Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 3812 ,
            USA
FEATURES             Location/Qualifiers
     source          1..2688
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            1..2688
                     /gene="ELF2"
     CDS             199..1980
                     /gene="ELF2"
                     /note="alternatively spliced; includes N-terminal
                     transactivation domain and 36nt insert not found in form
                     A1"
                     /codon_start=1
                     /product="ets family transcription factor ELF2A2"
                     /protein_id="AAF67192.1"
                     /db_xref="GI:7677534"
                     /translation="MASAVVDSGGSALELSDGGENQEGSTGLPTAVIVEEVESAR
                     LEQGYAAQVLVYDDETYMMQDVAEEQEVETENSETVEASVHSSNAHCTDKTIEAAEAL
                     LHMESPTCLRDSRSPVEVFVPPCISTPEFIHAAMRPDVITETVVEVSTEESEPMDASP
                     IPTSPDSHEPMKKKKVGRKPKTQQSPVSNGSPELGIKKKAREGKGNTTYLWEFLLDLL
                     QDKNTCPRYIKWTQREKGIFKLVDSKAVSKLWGKHKNKPDMNYETMGRALRYYYQRGI
                     LAKVEGQRLVYQFKDMPKNIVVIDDDKSETCNEDLAAAADDKSLERVSLSAESLLKAA
                     TAVRGGKNSSPLNCSRAEKGVARVVNITSPTHDGSSRSPTTTAPVSAAAAPRTVRVAM
                     QVPVVMTSLGQKISAVAVQSVNAGTGSPLITSTSPASASSPKVVIQTVPTVMPASTEN
                     GDRITMQPAKIITIPATQLAQCQLQAKSNLTGSGSINIVGTPLAVRALTPVSIAHGTP
                     VMRLSVPAQQASGQTPPRVISAVIKGPEVKSEAKKQEHDVKTLQLVEEKPADGNKTVT
                     HVVVVSAPSAIALPVTMKTEGLVTCEK"
     misc_feature    940..1068
                     /gene="ELF2"
                     /note="ets domain; DNA binding site"
     misc_feature    1510..1776
                     /gene="ELF2"
                     /note="LMO2-binding site; DNA binding site"
BASE COUNT      795 a    601 c    664 g    628 t
ORIGIN


alignment_scores:
            Quality:  276.50            Length:    369
              Ratio:  1.580               Gaps:     13
 Percent Similarity:  47.425   Percent Identity:  27.100

alignment_block:
 US-09-126-945B-2 x AF256217  ..

 Align seq 1/1  to: AF256217  from: 1  to: 2688

   4 AlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProPr 20
     :::|||   :::||||||||||||:::    :::||||||    ||||||
 138 TCGTCTGTGAGCCTTAGTTCCGTTGCTTCACAGCATTTGAGACTTCCTCA 187

  20 oAspThr........................ValSerArg.ThrGlyLeu 28
        |||                            |||  ::::::|||
 188 GGGAACGCCACGATGGCATCGGCGGTGGTTGACAGCGGAGGCTCTGCTCTG 237

  29 GluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerPr 45
     |||       :::   |||:::   :::   |||       |||   :::||
 238 GAGCTTCCTAGCGATGGAGGAGAAAATCAAGAGGGAGGTGACACGGGCCC 287

  45 oSer..........................ProProAlaThrProGluG 53
     |:::                           |||   |||       ||||
 288 TGACTGCCCGGCAGTGATTGTGGAGGAGGTTCCCAGTGCGAGGTTAGAGC 337

  53 lnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyr...... 67
     |||||   :::|||          |||   |||   |||       |||
 338 AGGGCTATGCTGCCCAGGTGCTGGTCTATGACGACGAGACTTATATGATG 387
```

```
  68 ..................................................Pr  68

 388 CAGGATGTGGCGGAAGAACAGGAAGTTGAGACGGAGAATTCAGAAACAGT 437

  68 oGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluG  89
       |||    |||      ||||||||||||      ||||||||      |
 438 GGAAGCATCAGTTCACAGCAGTAATGCTCACTGTACGGATAAGACCATTG 487

  85 IuProProGlu..............GluProGluGlnCysProValIle  96
     ||      |||                |||      |||    :::
 488 AAGCTGCTGAAGCCCTGCTTCATATGGAATCTCCTACCTGC...TTGAGG 534

  97 AspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuPh 118
     ||||||:::::|||:::              :::||||||      :::::
 535 GATTCAAGAAGTCCTGTGGAAGTG.....TTTGTCCCTCCTTGTATATC 578

 119 rLeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluV 140
     :      |||                 ::::::     ||:::   |||:
 579 AACTCCAGAGTTTATCCATGCTGCGATGAGGCCGGATGTCATCACAGAGA 628

 140 alLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAsp 146
     :::::  ::::::::|||                              ::::::
 629 CTGTACTGGAGGTGTCCACAGAA..................GAGTCTGAG 660

 147 ProMetAspTrpSerProSerAsnValGluLysTrpLeuLeuTrpThrGl 166
     |||||||||   ||||||
 661 CCGATGGATGCCTCGCCT.............................. 678

 166 uHisGluTyrArgLeuProProMetGlyLysAlaPheGluGluLeuAlaG 186
                ::|||          :::   :::   :::
 679 .............ATCCCTACTTCACCAGACAGCCACGAACCAATGAAAA 715

 180 lyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerPro 196
       |||:::        ::::::::::::         |||:::||||||
 716 AGAAAAAA.........GTTGGCCGTAAACCAAAGACCCAGCAGTCACCA 756

 197 LeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLysSerAlaAl 216
     :::::::::          :::|||   |||   ||:::::|||
 757 GTTTCCAATGGC......TCTCCTGAGTTAGGAATCAAGAAGAAAGCCAG 800

 217 aTrpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerT 236
           |||
 801 AGAAGGAAAA..................................... 810

 237 hrSerGluGlnSerTrpThrAspSerGluValAspSerCysSerSer... 246
                                                 |||
 811 ..............................................GGA 813

 247 GluProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHi 266
     :::      :::|||||||:::||||||   :::|||||      ::
 814 AACACAACCTATCTGTGGGAGTTTCTTTTAGATTTGCTTCAAGATAAAAA 863

 267 sSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePhet 286
     ::::        |||:::|||:::|||   ::::::||||||||||||||||
 864 TACTTGTCCCACGTATATTAAATGGACTCAGAGAGAAAACGCCATATTCA 913

 287 ysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsn 306
     |||::   ||||||     |||:::::::|||||||||   :::||||||
 914 AACTTGTGGACTCAAAGGCTGTCTCGAAGCTCTGGGGGAAGCATAAGAAC 963

 307 ArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyrTy 326
     :::|||   ||||||||||:::   ::::::|||:::::|||   ||||||
 964 AAACCAGATATGAACTATGAAACCATGGGGCGAGCGTTAAGATATTACTA 1013

 327 rLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrG 346
     |:::::|||||:::   |||   :::   :::|||||||||||||||||
1014 CCAAAGGGGAATTCTTGCAAAGGTTGAA...GGACAGAGGCTTGTATATC 1060

 347 InPhe 347
     |||||
1061 AGTTC 1065
```

```
seq_name: qb_htq13:AC055828

seq_documentation_block:
LOCUS       AC055828   177426 bp    DNA              HTG       14 MAY 2000
DEFINITION  Homo sapiens clone RP11 34016, WORKING DRAFT SEQUENCE, 46 unordered
            pieces.
ACCESSION   AC055828
VERSION     AC055828.2  GI:7801457
KEYWORDS    HTG; HTGS_PHASE1; HTGS DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177426)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11 34016
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177426)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslawsky,L., Bonkhalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Gindo,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Hanford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Joges,C., Kann,L., Karatas,A.,
            Klein,J., Lakospie,K., Lamazares,K., Landers,L., Lebovsky,J.,
            Levine,E., Liew,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,F.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Roquov,P., Rothman,D.,
            Roy,A., Santos,R., Schaner,S., Severy,P., Spencer,B.,
            Stange Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zaitoon,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18 APR 2000) Whitehead Institute/MIT Center for Genome
            ReSearch, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 14, 2000 this sequence version replaced gi:7582661.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            E:\tp...\ftp_pose.washington.edu_6M\RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
                Center project name: LB626
                Center clone name: 340_I_6
            ---------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 156302 bases at least Q40
                Consensus quality: 167006 bases at least Q30
                Consensus quality: 171400 bases at least Q20
                Insert size: 198000; agarose fp
                Insert size: 173926; sum of contigs
                Quality coverage: 4.5 in Q20 bases; agarose fp
                Quality coverage: 4.0 in Q20 bases; sum of contigs
            ---------- .
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 46 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
* be preserved.
*          1     1784: contig of 1784 bp in length
*       1785  1884: gap of       100 bp
*       1885     2989: contig of 1105 bp in length
*       2990  3089: gap of       100 bp
*       3090     5117: contig of 2028 bp in length
*       5118  5217: gap of       100 bp
*       5218     6626: contig of 1409 bp in length
*       6627  6726: gap of       100 bp
*       6727     8933: contig of 2207 bp in length
*       8934  9033: gap of       100 bp
*       9034    10648: contig of 1615 bp in length
*      10649 10748: gap of       100 bp
*      10749    12570: contig of 1822 bp in length
*      12571 12670: gap of       100 bp
*      12671    15798: contig of 3128 bp in length
*      15799 15898: gap of       100 bp
*      15899    19126: contig of 3228 bp in length
*      19127 19226: gap of       100 bp
*      19227    22308: contig of 3082 bp in length
*      22309 22408: gap of       100 bp
*      22409    25462: contig of 3054 bp in length
*      25463 25562: gap of       100 bp
*      25563    29758: contig of 4196 bp in length
*      29759 29858: gap of       100 bp
*      29859    32899: contig of 3041 bp in length
*      32900 32999: gap of       100 bp
*      33000    37286: contig of 4287 bp in length
*      37287 37386: gap of       100 bp
*      37387    41144: contig of 3758 bp in length
*      41145 41244: gap of       100 bp
*      41245    46573: contig of 5329 bp in length
*      46574 46673: gap of       100 bp
*      46674    50564: contig of 3891 bp in length
*      50565 50664: gap of       100 bp
*      50665    55581: contig of 4917 bp in length
*      55582 55681: gap of       100 bp
*      55682    60807: contig of 5126 bp in length
*      60808 60907: gap of       100 bp
*      60908    65566: contig of 4659 bp in length
*      65567 65666: gap of       100 bp
*      65667    69323: contig of 3657 bp in length
*      69324 69423: gap of       100 bp
*      69424    73505: contig of 4082 bp in length
*      73506 73605: gap of       100 bp
*      73606    79191: contig of 5586 bp in length
*      79192 79291: gap of       100 bp
*      79292    84609: contig of 5318 bp in length
*      84610 84709: gap of       100 bp
*      84710    90446: contig of 5737 bp in length
*      90447 90546: gap of       100 bp
*      90547    96382: contig of 5836 bp in length
*      96383 96482: gap of       100 bp
*      96483   101634: contig of 5152 bp in length
*     101635 101734: gap of       100 bp
*     101735   108444: contig of 6710 bp in length
*     108445 108544: gap of       100 bp
*     108545   115975: contig of 7431 bp in length
*     115976 116075: gap of       100 bp
*     116076   122169: contig of 6094 bp in length
*     122170 122269: gap of       100 bp
*     122270   130312: contig of 8043 bp in length
*     130313 130412: gap of       100 bp
*     130413   137841: contig of 7429 bp in length
*     137842 137941: gap of       100 bp
*     137942   148046: contig of 10105 bp in length
*     148047 148146: gap of       100 bp
*     148147   157364: contig of 9218 bp in length
*     157365 157464: gap of       100 bp
*     157465   167083: contig of 9619 bp in length
*     167084 167183: gap of       100 bp
*     167184   177426: contig of 10243 bp in length.
FEATURES             Location/Qualifiers
     source          1. .177426
                     /organism "Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone "RP11-34O16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1784
                     /note="assembly_fragment"
     misc_feature    1885. .2989
                     /note="assembly_fragment"
     misc_feature    3090. .5117
                     /note "assembly_fragment"
     misc_feature    5218. .6626
                     /note "assembly_fragment"
     misc_feature    6727. .8933
                     /note "assembly_fragment"
     misc_feature    9034. .10648
                     /note "assembly_fragment"
     misc_feature    10749. .12570
                     /note "assembly_fragment"
     misc_feature    12671. .15798
                     /note "assembly_fragment"
     misc_feature    15899. .19126
                     /note "assembly_fragment"
     misc_feature    19227. .22308
                     /note="assembly_fragment"
     misc_feature    22409. .25462
                     /note "assembly_fragment"
     misc_feature    25563. .29758
                     /note "assembly_fragment"
     misc_feature    29859. .32899
                     /note "assembly_fragment"
     misc_feature    33000. .37286
                     /note "assembly_fragment"
     misc_feature    37387. .41144
                     /note "assembly_fragment"
     misc_feature    41245. .46573
                     /note "assembly_fragment"
     misc_feature    46674. .50564
                     /note "assembly_fragment"
     misc_feature    50665. .55581
                     /note "assembly_fragment"
     misc_feature    55682. .60807
                     /note "assembly_fragment"
     misc_feature    60908. .65566
                     /note "assembly_fragment"
     misc_feature    65667. .69323
                     /note "assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    69424. .73505
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    74606. .79191
                     /note "assembly_fragment"
     misc_feature    79292. .84609
                     /note "assembly_fragment"

alignment_scores:
            Quality:  276.50              Length:     305
              Ratio:  1.589                 Gaps:       9
 Percent Similarity:  57.049   Percent Identity: 27.213

alignment_block:
 US-09-126-945B-2 x AC055828   ..

 Align seq 1/1  to: AC055828  from: 1  to: 177426

      60 LeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLy 76
         ||:::  :::::::::::  |||  ||:::     ::::::
   33507 TTAAACATCCTTGAAATTCTTAAACCTGCGGACACTGTTCAGCTTGTTAT 33556

      76 sAlaProGlyAlaSerSerArgGluGluProProGluGluProGluGlnC 93
```

```
                 |||    |||          |||:::      :::       |||:::
  33557 TCATCCAGATGCCCACCATGCTGAATCAGAAGCACATCTTGTTCAAGAAG 33606

     93 ysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValPro 109
        ::   ||||||                 :::|||||   :::
  33607 CTCAAGTGATA.....................ACTCTTGATGGCACAAAA 33645

    110 GlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetVa 126
                |||:::::::::::::::::   ||||||||           ::
  33646 CACATCACAACCATTTCAGATGAAACTTCAGAACAAGTGACAAGATGGGC 33695

    126 lValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnI 143
        ::::      :::|||:::      :::          :::   |||:::|
  33696 TGCT.....GCACTGCAAGGCTATAGGAAAGAACAAGAACGCCTTGGCA 33729

    143 IleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeu 159
        ||       ||||||:::::::|||||    :::::|||      |||:::
  33730 TACCCTATGATCCCATACAGTGGTCCACAGACCAAGTCCTGCATTGGGTC 33779

    160 LeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGl 176
        :::|||:::      :::::   :::       :::       :::
  33780 GTTTGGCTAATGAAGGAATTCAGCATGACCGATATAGACCTCACCACACT 33829

    176 nGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgS 193
        :::::::::|||:::|||||||||:::::::::::|||:::|||   |
  33830 CAACATTTCGGGGAGAGAATTATGTAGTCTCAACCAAGAAGATTTTTTC 33879

    193 lnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrp 209
        |||||   |||      |||:::::|||   :::|||||:::::
  33880 AGCGGGTTCCT...CAGGGAGAAATTCTCTGGAGTCATCTGGAACTTCTC 33926

    210 LysSerAlaAlaTrpMetLysGluArgThrSer................. 220
        ::::::   :::       :::::::::   :::
  33927 CGAAAATATGTATTGCAAGTCAAGAACAACAGATGAATGAAATAGTTAC 33976

    221 .........................ProGlyAlaIleHisTyrCysAlaS 239
                                 |||:::::::::::::   ::::::
  33977 AATTGATCAACCTGTGCAAATTATTCCAGCATCAGTGCAATCTGCTACAC 34026

    229 erThrSerGluGluSerTrpThrAspSer..................... 238
          |||:::   :::       :::::|||
  34027 CTACTACCATTAAAGTTATAAATAGTAGTGTGAAGCCAGCCAAAGTACAA 34076

    239 ................GluValAspSerSerCysSerGlyGln........ 247
                       :::   |||   |||      |||:::
  34077 AGACGCCCAGGATTTCAGAAGATAGAAGCTCACCTGGGAACAGAACAGC 34126

    248 ..........ProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuL 261
                 |||:::|||   ||||||||   ||||||||
  34127 AAACAATGGCCAAATCCAACTATGACAGTTTTTGCTAGAACTTCTTACTG 34176

    261 ysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGly 277
              :::            |||   |||:::::   :::||||||
  34177 ATAAGGACGCTCGAGAC...TGTATTTCTTGGGTTGGT...GATAAAGGT 34220

    278 IlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleAs 294
          |||||:::::::::          |||||:::   ||||||   ||
  34221 GAATTTAAGCTAAATCAGCCTGAACTGGTTCGCACAAAAATGGGGACAGAG 34270

    294 qLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgS 311
        |||||||:::|||:::||||||||||:::||||||||||||:::::|||
  34271 TAAAAATAAGCCTACGATGAACTATGACAAACTCAGTCGTGCATTAAGAT 34320

    311 InTyrTyrLysLysGlyIleIleArgLysProAspIleSerGluArgLeu 327
          ||||||         :::|||   |||   :::   :::::|||:::
  34321 ATTATTATGATGGGGACATGATTTGTAAAGTTCAA...GGCAAGAGATTT 34367

    328 ValTyrGlnPheVal 332
        ||||||:::||||||
  34368 GTGTACAAGTTTGTC 34382

seq name: qb rol:AF256216
```

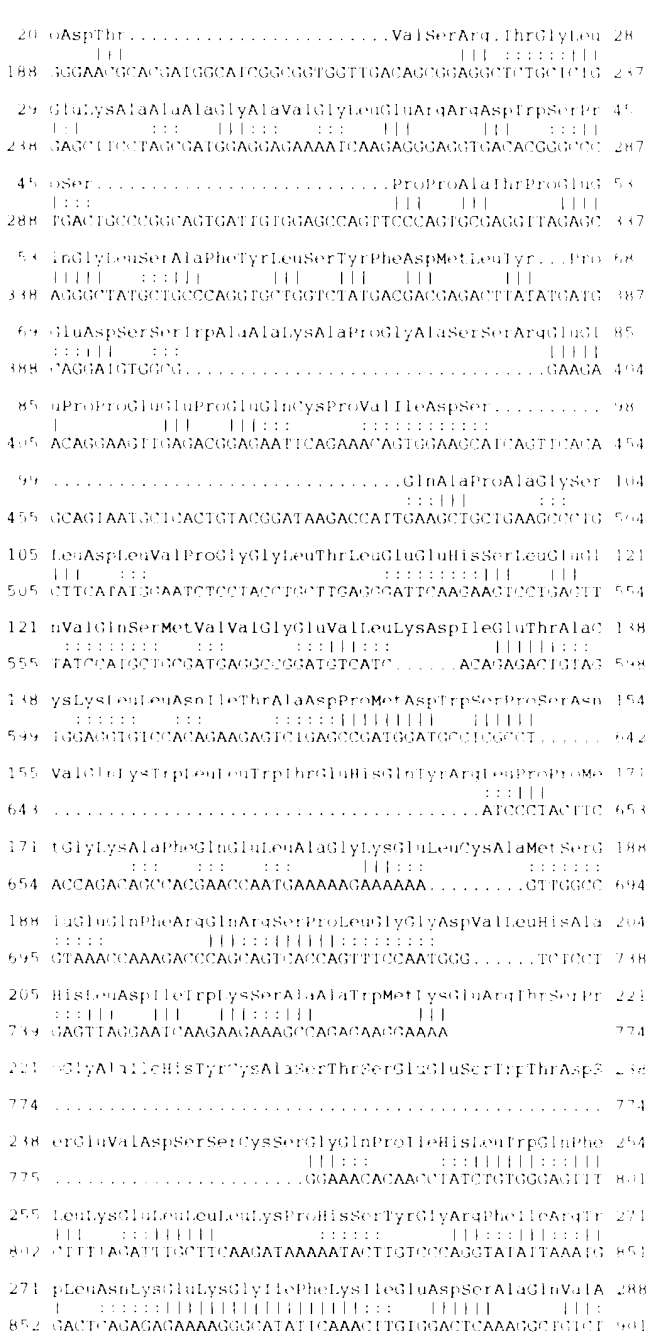
```
seq documentation block:
LOCUS       AF256216      2652 bp    mRNA            ROD       02-MAY-2000
DEFINITION  Mus musculus ets family transcription factor ELF2A1 (ELF2) mRNA,
            complete cds, alternatively spliced.
ACCESSION   AF256216
VERSION     AF256216.1  GI:7677530
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2652)
  AUTHORS   Wilkinson,D.A., Neale,G.A., Mao,S., Naeve,C.W. and Goorha,R.M.
  TITLE     Elf 2, a rhombotin 2 binding ets transcription factor: discovery
            and potential role in T cell leukemia
  JOURNAL   Leukemia 11 (1), 86-96 (1997)
  MEDLINE   97154704
REFERENCE   2  (bases 1 to 2652)
  AUTHORS   Wilkinson,D.A., Neale,G.A.M., Mao,S., Fernandes,E.R.,
            Davenport,J.W., Naeve,C.W. and Goorha,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR 2000) Virology and Molecular Biology, St. Jude
            Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38120,
            USA
FEATURES             Location/Qualifiers
     source          1. .2652
                     /organism "Mus musculus"
                     /db_xref "taxon:10090"
     gene            1. .2652
                     /gene "ELF2"
     CDS             199. .1944
                     /gene "ELF2"
                     /note "alternatively spliced; includes N terminal
                     transactivation domain; lacks 36nt insert found in
                     form Az"
                     /codon_start 1
                     /product "ets family transcription factor ELF2A1"
                     /protein_id "AAF67191.1"
                     /db_xref "GI:7677531"
                     /translation "MASAVVDSGGSALELPSDGGENQEGGDTGPDCPAVTVEPVPSAR
                     LEQGYAAQVLVYDDETYMMQDVAEEQEVETENSETVEASVHSSNAIHCTDKTIEAAEAL
                     LHMESPTCLRDSRSPEFIHAAMRPDVTTETVVEVSTEESEPMDASPTPTSPDSHEPMK
                     KFEVGPEPMTQQSPVSNGSPELCPKKAPEPGNETYLWSFLLTLQDKNICPYLNW
                     TQREFCTEELVDSFAVSELWGEHENEPDMNYEIMCRALEYYYQRGILAKVEGQRLVYQ
                     FKDMPKNIVVIDDDKSETCPEDLAAAADDKSLERVSLSAESLLKAATAVRGGKNSSPL
                     NCSKAEKGVARVVNITSPTHDGSSRSPTTTAPVSAAAAPRTVRVAMQVPVVMTSLGQK
                     ISAVAVQSVNAGTGSPLTTSTSPASASSPKVVIQTVIPTVMPASTENGDRTIMQPAKTI
                     IIPAIQLAQCQLQAKSNIPGSGSINIVGIPLAVRALTPVSIAHGTPVMRLSVPAQQAS
                     GQTPPRVISALLPGPGESEAEPQEHDVEIIGEVEESGALPGNKIVIHVVVVSAPSAIA
                     LPVIMKIEGLVICEK"
     misc_feature    904. .1062
                     /gene "ELF2"
                     /note "ets domain; DNA binding site"
     misc_feature    1474. .1740
                     /gene "ELF2"
                     /note "LMO2-binding site; DNA binding site"
BASE COUNT      788 a    592 c    657 g    615 t
ORIGIN


alignment scores:
            Quality:  275.50              Length:     461
              Ratio:  1.592                 Gaps:      12
 Percent Similarity:  47.922   Percent Identity: 26.593

alignment block:
 US-09-126-945B-2 x AF256216   ..

 Align seq 1/1  to: AF256216  from: 1  to: 2652

        4 AlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProPt 20
          ::|||   :::|||||||||||:::   :::|||||   ||||||
      138 TCGTCTGTGAGCCTTAGTTCCGTTGCTTCACAGCATTTGAGACTTCCTCA 187
```



```
seq_name: qb_ov:XLFLIG

seq_documentation_block:
LOCUS       XLFLIG       4403 bp    mRNA            VRT       21-JUN-1994
DEFINITION  X.laevis mRNA XLFLI.
ACCESSION   X66979
VERSION     X66979.1  GI:505486
KEYWORDS    fli gene; Fli protein.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 4403)
  AUTHORS   Wolff,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-1992) C.M. Wolff, Institut de Biologie
            Moleculaire et Cellulaire du C N R S, 15 Rue Rene Descartes, 67084
            Strasbourg Cedex, FRANCE
REFERENCE   2  (bases 1 to 4403)
  AUTHORS   Meyer,D., Wolff,C.M., Stiegler,P., Senan,F., Befort,N., Befort,J.J.
            and Remy,P.
  TITLE     Xl-fli, the Xenopus homologue of the fli-1 gene, is expressed
            during embryogenesis in a restricted pattern evocative of neural
            crest cell distribution
  JOURNAL   Mech. Dev. 44 (2-3), 109-121 (1993)
  MEDLINE   94206844
FEATURES             Location/Qualifiers
     source          1. .4403
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /dev_stage="unfertilized eggs"
                     /clone_lib="cDNA in lambda gt10"
     gene            100. .1461
                     /gene="Xlfli"
     CDS             100. .1461
                     /gene="Xlfli"
                     /codon_start=1
                     /product="XLFli protein"
                     /protein_id "CAA47489.1"
                     /db_xref="GI:505487"
                     /db_xref="SWISS-PROT:P41157"
                     /translation "MDGLIKEALSVVSDDQSLFDSAYGASSHLSKADMTASANPDYGQ
                     PHKINPIPPQQHWINQPMHVNTKPEYFHMNGSPFSPVDCSINKCSKLIGGSEGNAMIY
                     TYMDEKNGPPPPNMTTNERRVIVPADPALWSQDHVPQWLEWAIKEYGIVFIDCSLPQN
                     IEGKELCKMSKELIESISIYNIEVCLSHLNYLKLKSSSSLGYNIQADIIGLSSHLTAFE
                     DPSYEAVRPSGWGNSMSSPVTKSPPMGGTQNVNKSGDQQRSQPDPYQIIGPTSSKIAN
                     PGSGQIQLWQFLLELLSDSSNASCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYD
                     KLSRALRYYYDKSIMKKVHGKRYAYKFDFHGIAQALQPHPTESSMYKYPSDISYMPSY
                     HSHQQKVNFVPSHPSSMPVTSSGFFGATSPYWNSPSANIYPNPNVPRHPNTHVQSHIG
                     GPY"
     misc_feature    931. .1185
                     /gene="Xlfli"
                     /note="ets-domain"
     polyA_signal    4449. .4455
BASE COUNT     1382 a    846 c    871 g   1314 t
ORIGIN


alignment_scores:
          Quality:  275.00          Length:   241
```

```
            Ratio:  2.083                 Gaps:      9
 Percent Similarity:  54.772   Percent Identity:  32.365

alignment block:
 US-09-126-945B-2 x XLFLIG   ..

 Align seq 1/1   to: XLFLIG   from: 1   to: 4403

     143 IleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLe 189
         :::   |||||||||       ||||||   :::::::|||:::::|||||
     472 GTACCTGCACATCCCGCTTGTGGACGCCAGGACCACGTACGCCAGTGGCT 521

     159 uLeuTrpThrGluHisGlnTyrArgLeuProProMetGly...LysAlaP 175
         |   |||:::      :::|||   |||       :::      :::   |
     522 GGAGTGGGCAATTAAGGAATATGGACTTGTGGAAATTGACTGCTCCCTCT 571

     175 heGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPhe 191
         |||||:::::   ||||||||||||||   ||||||:::|||:::|||
     572 TTCAGAATATTGATGGCAAGGAACTGTGTAAAATGAGCAAAGAGGATTTC 621

     192 ...ArqGluArqSerProLeuGlyGlyAspValLeuHisAlaHisLeuAs 207
             |||:::   |||      :::   :::||||||   :::|||||||:
     622 CTGCGGAGCACATCCATATATAACACAGAGGTTTTACTTTCTCATCTCAA 671

     207 pIleTrpLysSerAlaAla..............................  213
         :   ::::::::::::
     672 CTACCTCAGGGATAGTAGCTCATCATTCGGTTATAACACTCAGGCACATA 721

     213 ..............................................  213

     722 CAGACCAATCATCGCGTTTCACACCCAAACAGGATCCCTCCTATGAACCG 771

     214 ..............TrpMetLysGluArqThrSerPro.........  223
                       |||   ::::::   :::|||||
     772 GTAACGAGATCTGGATGGGCAAATTCCATGAGCTCACCTGTCACAAAAAG 821

     222 ..........GlyAlaIleHisTyrCysAlaSerThrSerGluGluSerT 235
                  |||:::   :::          :::::::::::::::::::
     822 GCCTCCCATGGGAGGAACCCAAAATCTGAATAAAAGCGGGGATCAGCAAC 871

     235 rp..............................ThrAspSerGluVal  240
                                        |||:::|||:::::
     872 GTTCTCAGCCAGATCCATACCAAATATTAGGGCCCACAACGCAGCAGTG 921

     241 AspSerSerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysC 257
         :::       |||||||||   |||:::|||||||||||||||||   ||
     922 GCTAACCCAGGGAGTGGCCAA...ATTCAGCTGTGGCAGTTTCTCTTGGA 968

     257 uLeuLeuLeuLysProHisSerTyrGlyArqPheIleArqTrpLeuAsnL 274
         |   |||||||:::      |||  :::      |||  |||   :::
     969 G...TTGCTATCTGACAGTTCAAACGCTAGTTGTATCACATGG...GAAG 1012

     274 ysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArqLeu 290
         :::|||   |||||||:::  |||      :::|||||||||
    1013 GCACTAATGGAGAGTTCAAAATGACCGACCCAGATGAAGTAGCGAGACGT 1062

     291 TrpGlyIleArqLysAsnArqProAlaMetAsnTyrAspLysLeuSerAr 307
         ||||||  ||||||:::::::|||  ||||||||||||||||||||||||
    1063 TGGGGTGAAAGGAAAAGCAAGCCTAACATGAACTATGACAAATTAAGCGG 1112

     307 qSerIleArqGlnTyrTyrLysLysGlyIleIleArqLysProAspIleS 324
         |:::::::||   ||||||   |||:::|||:::   |||         :
    1113 AGCACTAAGATACTATTATGATAAAAGTATAATGACTAAAGTCCATC...G 1159

     324 erGlnArqLeuValTyrGlnPhe 338
         :::::|||   :::|||:::|||
    1160 GTAAACGTTATGCCTACAAATTT 1182

seq name: qb ov:CCRNAFLI

seq documentation block:
LOCUS       CCRNAFLI     3490 bp    mRNA            VRT       07-JUL-1998
DEFINITION  Coturnix coturnix mRNA for transcription factor FLI, clone 10.1.

ACCESSION   Y14774
VERSION     Y14774.1  GI:3269302
KEYWORDS    fli gene; FLI oncoprotein; transcription factor.
SOURCE      common quail.
  ORGANISM  Coturnix coturnix
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Coturnix.
REFERENCE   1  (bases 1 to 3490)
  AUTHORS   Mager,A.M., Grapin Botton,A., Ladjali,K., Meyer,D., Wolff,C.M.,
            Stiegler,P., Bonnin,M.A. and Remy,P.
  TITLE     The avian fli gene is specifically expressed during embryogenesis
            in a subset of neural crest cells giving rise to mesenchyme
  JOURNAL   Int. J. Dev. Biol. 42 (4), 561-572 (1998)
  MEDLINE   98358003
REFERENCE   2  (bases 1 to 3490)
  AUTHORS   Stiegler,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1997) P. Stiegler, Centre National de la
            Recherche, Scientifique, U.P.R. 9005 du C.N.R.S, 15 Rue Rene
            Descartes, 67084 Strasbourg Cedex, FRANCE
COMMENT     Related sequence Y14774.
FEATURES             Location/Qualifiers
     source          1..3490
                     /organism="Coturnix coturnix"
                     /db_xref="taxon:9091"
                     /chromosome="1"
                     /dev_stage="4 day quail embryo"
                     /clone_lib="pCbNA1"
                     /clone="10.1"
                     /map="q41 33"
     gene            215..1518
                     /gene="fli"
     CDS             215..1518
                     /gene="fli"
                     /function="transcription factor"
                     /codon_start=1
                     /product="FLI transcription factor"
                     /protein_id="CAA75077.1"
                     /db_xref="GI:3269303"
                     /db_xref="SPTREMBL:O93425"
                     /translation="MDGTIKEALSVVSEDQSLFDSTYGAATHLPKADMTASGNPEYGQ
                     PHKINPLPPQQEWINQPVRVNVKREEYDPMNGSPEESPVDCSVNKCSKLVGAGIESNPMS
                     YSTYMDEKNGPPPNMTTNERRVIVPADPTLWTQEHVRQWLEWAIKEYGLMEIDTTFFQ
                     NMDGKELCKMNKEDFLRTTSLYNTEVLLSHLSYLRESSSLLAYNTPSHTFASSRLATK
                     EGPPVAGTQNVNKTTEQQRPQPDPYQTLGPTSSRLANPGSGQIQLWQFLLELLSDSSN
                     ASCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGK
                     RYAYKFDFHGIAQALQPHPTESSMYKYPSDLSYMPSYHAHQQKVNFVPPHPSSMPVTS
                     SSFFGAASPYWTSPAGSTYPNPNVPRHPNAHVSPHLGSYY"
BASE COUNT      1026 a    770 c    752 g    942 t
ORIGIN


alignment_scores:
            Quality:  274.00              Length:    221
              Ratio:  2.124                 Gaps:      9
 Percent Similarity:  58.471   Percent Identity:  34.842

alignment_block:
 US-09-126-945B-2 x CCRNAFLI   ..

 Align seq 1/1   to: CCRNAFLI   from: 1   to: 3490

     143 IleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLe 189
         :::   |||||||||       ||||::   :::::::|||:::::|||||
     590 GTCCCAGCAGATCCCACCCTATGGACCCAGGAACACGTGCGCCAGTGGCT 639

     189 uLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAla...P 175
         |   |||:::      :::|||   |||      :::      :::   |
     640 GGAATGGGCCATAAAGGAGTATGGATTAATGGAGATTGACACCACCTTCT 689

     175 heGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPhe 191
         |||||:::::   ||||||||||||||   |||:::::::::::|||
```

```
     690 ICCAGAATATGGATGGCAAAGAGCTCTGCAAAATGAACAAGGATGACTTC 739

     192 ...ArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAs 237
           |||      |||      ...   ...||||||   ...|||||...
     740 CTCCGAACCAGCTGCCTCTACAACAGAAGTTCTGTTGTCTCACCTCAG 789

     207 pIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAla1 224
           :     ::::::::::::   :::   ::::::|||:::
     790 TTACCTCAGGGAAAGTAGCTCACTGCTGGCCTACAACACTCCATCC... 835

     224 leHisTyrCysAlaSer..................................... 229
           |||      ||||||
     836 ..CACACAGAAGCTTCCTCACGTCTCGCCACCAAAGAAGGTCCTCCTGTT 885

     230 ........................ThrSerGluGluSerTrp........ 245
                                  |||:::|||:::::
     884 GCAGGGACACAAAATGTGAACAAGACAACAGAACAGCAACGGCCCCAACC 933

     246 ........................ThrAspSerGluValAspSerSerC 261
                                  |||:::|||:::::   :::
     934 AGATCCATATCAAATCCTGGGGCCCACCAGTAGTCGTCTTGCCAATCCTG 983

     262 ysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeu 269
           |||||||||   |||:::|||||||||||||||||   |||   ||||||
     984 GGAGTGGCCAG   ATACAACTGTGGCAGTTCCTCCTCGAG   TTGCTG 1027

     261 LysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysL 277
           :::   |||   :::   |||   |||   :::   :::|||
    1028 TCGGACAGTTCCAATGCCAGCTGTATCACACATGG...GAAGGGACCAATGG 1074

     277 yIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleA 294
           |   ||||||:::   |||      :::||||||||||   ||||||   |
    1075 GGAATTCAAGATGACAGACCCACATGAAGTGGCACGGCGCTGGGGAGAAC 1124

     294 rgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArg 310
           |||||:::::::|||   |||||||||||||||||||||||||:::|||
    1125 GCAAAAGCAACGCCAACATGAATTATGACAAGCTGAGCCGAGCGCTTCGA 1174

     311 GlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLe 327
           ||||||   |||:::|||:::   |||         :::::|||
    1175 TACTACTATGATAAGAACATTATGACCAAAGTGCAT...GGCAAAAGGTA 1221

     327 uValTyrGlnPhe 331
           :::|||:::|||
    1222 TGCCTACAAATTT 1234

seq_name: qb_ov:CCFLIONCO

seq_documentation_block:
LOCUS       CCFLIONCO    3545 bp    mRNA            VRT       07-JUL-1998
DEFINITION  Coturnix coturnix mRNA for transcription factor FLI, clone 6.1.
ACCESSION   Y14774
VERSION     Y14774.1  GI:3269304
KEYWORDS    fli gene; FLI oncoprotein; transcription factor.
SOURCE      common quail.
  ORGANISM  Coturnix coturnix
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Coturnix.
REFERENCE   1  (bases 1 to 3545)
  AUTHORS   Stiegler,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-1997) P. Stiegler, Centre National de la
            Recherche Scientifique, U.P.R. 9005 du C.N.R.S, 15 Rue Rene
            Descartes, 67084 Strasbourg Cedex, FRANCE
REFERENCE   2  (bases 1 to 3545)
  AUTHORS   Mager,A.M., Grapin-Botton,A., Ladjali,K., Meyer,D., Wolff,C.M.,
            Stiegler,P., Bonnin,M.A. and Remy,P.
  TITLE     The avian fli gene is specifically expressed during embryogenesis
            in a subset of neural crest cells giving rise to mesenchyme
  JOURNAL   Int. J. Dev. Biol. 42 (4), 561-572 (1998)
  MEDLINE   98458003
COMMENT     Related sequence Y14773.

FEATURES             Location/Qualifiers
     source          1. .3545
                     /organism="Coturnix coturnix"
                     /db_xref="taxon:9091"
                     /chromosome="1"
                     /dev_stage "4 day quail embryo"
                     /clone_lib="pCDNA1"
                     /clone "6.1"
                     /map="q41-44"
     gene            396. .1595
                     /gene="fli"
     CDS             396. .1595
                     /gene "fli"
                     /function="transcription factor"
                     /codon_start 1
                     /product="FLI transcription factor"
                     /protein_id="CAA75078.1"
                     /db_xref="GI:3269305"
                     /translation "MTASGNETYGQPHEINPLFIQQEWTNPIVEVNVEKETDHMMLSP
                     ESPVDCSVNKCSKLVGAGTESNPMSYSTYMDEKNGPPPNMTTNERRVIVPADPTLWTQ
                     EHVPQWFWATKEYGLMFKLDTTFFQNMDGKFLCKMNKHDFLRTTSLYNPEVLILSHLSY
                     LREESSLLAYNTPSHTEASSRLATKEGPPVAGTQNVNKTTEQQRPQPDPYQILGPTSS
                     RLANPGSGQIQLWQFLLELLSDSSNASCITWEGTNGEFKMTDPDEVARRWGERKSKPN
                     MNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPTESSMYKYPSDLSY
                     MPSYHAHQQKVNFVPPHPSSMPVTSSSSFFGAASPYWTSPAGSTYDNPNVPEHPNAHVS
                     PHLGSYY"
     misc_feature    1071. .1820
                     /gene="fli"
                     /note "ETS binding domain"
BASE COUNT     1002 a    803 c    768 g    972 t
ORIGIN


alignment_scores:
           Quality:  274.00            Length:    221
             Ratio:  2.124               Gaps:      9
 Percent Similarity:  58.371   Percent Identity: 34.842

alignment_block:
 US-09-126 945B 2 x CCFLIONCO  ..

 Align seq 1/1  to: CCFLIONCO  from: 1  to: 3545

     143 IleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysIrpLe 159
           :::   |||||||||      |||:::   :::::::|||:::::|||||
     672 GTCCCAGCAGATCCCACCCTATGGACCCAGGAACACGTGCGCCAGTGGCT 721

     159 uLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAla...P 175
           |   |||:::      :::|||   |||      :::      :::   |
     722 GGAATGGGCCAIAAAGGAGIATGGAIIAATGGACATTGACACCACCTICT 771

     175 heGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPhe 191
           ||||::::::   ||||||||||||||||   |||:::::::::::|||
     772 TCCAGAATATGGATGGCAAAGAGCTCTGCAAAATGAACAAGGATGACTTC 821

     192 ...ArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeuAs 207
           |||      |||      :::   :::||||||   :::|||||||::
     822 CTCCGAACCAGCTGCCTCTACAACAGAAGTTCTGTTGTCTCACCTCAG 871

     207 pIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerProGlyAla1 224
           :     ::::::::::::   :::         ::::::|||:::
     872 TTACCTCAGGGAAAGTAGCTCACTGCTGGCCTACAACACTCCATCC.... 917

     224 leHisTyrCysAlaSer..................................... 229
           |||      ||||||
     918 ..CACACAGAAGCTTCCTCACGTCTCGCCACCAAAGAAGGTCCTCCTGTT 967

     230 ........................ThrSerGluGluSerTrp........ 245
                                  |||:::|||:::::
     966 GCAGGGACACAAAATGTGAACAAGACAACAGAACAGCAACGGCCCCAACC 1015

     246 ........................ThrAspSerGluValAspSerSerC 244
                                  |||:::|||:::::   :::
```

```
     1016 AGATCCATATCAAATCCTGGGGCCCACCAGTAGTCGTCTTGCCAATCCTG 1065

      244 ysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeu 293
            ||||||||||   |||:::|||||||||||||||||   |||   ||||||
     1066 GGAGTGGGCAG...ATACAACTGTGGCAGTTCCTCCTCGAG...TTGCTG 1109

      261 LysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGl 272
          :::      |||   :::      |||   |||   :::       :::||
     1110 TCGGACAGTTCCAATGCCAGCTGTATCACATGG...GAAGGGACCAATGG 1156

      277 yIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleA 294
          |   ||||||||:::   |||       :::||||||||||   ||||||   |
     1157 GGAATTCAAGATGACAGACCCAGATGAAGTGGCACGGCGCTGGGGAGAAC 1206

      294 rgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArg 310
          ||||||:::::|||   ||||||||||||||||||||||||||:::::|||
     1207 GCAAAAGCAAGCCCAACATGAATTATGACAAGCTGAGCCGAGCCCTTCGA 1256

      311 GlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLe 327
            ||||||   |||:::|||:::   |||         :::::|||
     1257 TACTACTACGATAAGAACATTATGACCAAAGTGCAT...GGCAAAAGGTA 1303

      327 uValTyrGlnPhe 331
            :::|||:::|||
     1304 TGCCTACAAATTT 1316

Seq name: qb rol:MMFLII

Seq documentation block:
LOCUS       MMFLII       1729 bp    mRNA            ROD       02-SEP 1991
DEFINITION  Mouse Fli 1 mRNA for retroviral integration site.
ACCESSION   X59421
VERSION     X59421.1  GI:50974
KEYWORDS    c ets 1 gene; ets gene family; fli-1 gene; integration site.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuroqnathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1729)
  AUTHORS   Letwin,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08 MAY 1991) K. Letwin, Samuel Lunenfeld Res Inst at
            Mount Sinai Hospital, Div of Mol and Developmental Biol, 600
            University Avenue, Toronto Ontario M5G 1X5, CANADA
REFERENCE   2  (bases 1 to 1729)
  AUTHORS   Ben David,Y., Giddens,E.B., Letwin,K. and Bernstein,A.
  TITLE     Erythroleukemia induction by Friend murine leukemia virus:
            insertional activation of a new member of the ets gene family,
            Fli 1, closely linked to c ets 1
  JOURNAL   Genes Dev. 5 (6), 908 918 (1991)
  MEDLINE   91257578
FEATURES             Location/Qualifiers
     source          1. .1729
                     /organism "Mus musculus"
                     /strain "balb/c"
                     /db xref "taxon:10090"
                     /germline
                     /cell type "Friend erythroleukemia"
                     /cell line "CB7"
     mRNA            1. .1729
                     /gene "Fli-1"
                     /evidence experimental
     gene            1. .1729
                     /gene "Fli 1"
     CDS             245. .1603
                     /gene "Fli 1"
                     /codon start 1
                     /product "retrovirus integration site"
                     /protein id "CAA42056.1"
                     /db xref "GI:50975"
                     /db xref "MGD:MGI:95554"
                     /db xref "SWISS-PROT:P26323"
                     /translation "MDGLTKEALSVVSEDQSLFESAYGAAAHVSAADMIASGSPERGG
                     PHKINPLPPQQEWINQPVRVNVKREYDHMNGSRESPVDCSVSKCNKLVGGGEANPMNY
                     NSYMDEKNGPPPPNMTTNERRVIVPADPTLWTQEHVRQWLEWAIKEYGLMEIDTSFFQ
                     NMDGKDLCKMTKEDFQRLVPSFPSSMPVTSPSHFTGAASSQYWTSPTAGIYPNPSVPR
                     HPNTHVDPSHLGSDPLNVFE
                     DPSYDSVRRGAWNNNMNSGINKSPLLGGSQTMGKNTFQRPQPDPYQTLGPTSSKLANP
                     GSGRQTQLWQFLLELLSDSANASCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDK
                     LSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPTETSMYKYPSDISYMPSYH
                     AHQQKVNFVPSHPSSMPVTSPSTFGAASSQYWTSPTAGIYPNPSVPRHPNTHVPSHLGS
                     YY"
BASE COUNT      466 a    484 c    436 g    343 t
ORIGIN


alignment_scores:
            Quality:  272.50              Length:    328
              Ratio:  1.703                 Gaps:     13
 Percent Similarity:  48.780   Percent Identity: 28.354

alignment_block:
 US-09-126-945B-2 x MMFLII   ..

 Align seq 1/1  to: MMFLII  from: 1  to: 1729

       80 AlaSerSerArgGluGluProProGluGluProGlu.....GlnCysPr 94
          |||:::   :::::::::::|||||                  :::   ||
      373 GCAGCTCCACAAAATCAACCCCCTGCCACCGCAGCAGGAGTGGATCAACC 422

       94 oValIleAspSerGlnAlaProAlaGlySer.............LeuA 106
          |:::   :::||||||||::   ||||||                ::::
      423 AGCCAGTGAGAGTCAATGTCAAGCGGGAGTATGACCACATGAATGGATCC 472

      106 spLeuValProGlyGlyLeuThrLeuGluGlu.............. 116
          ::   |||   |||||||||       ::::::
      473 AGGGAGTCTCCGGTGGACTGCAGTGTCAGCAAATGTAACAAGCTGGTGGG 522

      117 ................HisSerLeu.GluGlnValGlnSerMetValV 127
                          |||:::||| ::::::   ::::::
      523 CGGAGGCGAAGCCAACCCCATGAACTATAATAGCTACATGGATGAGAACA 572

      127 alGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIle 143
            |||           ::::::   |||      :::   :::   :::
      573 ACGGCCCCCCTCCTCCCAACATGACCACCAACGAACGGAGAGTCATTGTG 622

      144 ThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLe 160
          |||||||||      |||:::  ::::::|||:::::|||||
      623 CGTGCAGACCCCACACTGTCGACACAGGAGCACGTTCGACACTGCCTCCA 672

      160 uTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAla...PheG 176
            |||:::     :::|||   |||      :::   :::   ||||
      673 GTGGGCTATAAAGGAATACGGATTGATGGAGATTGACACTTCCTTCTTCC 722

      176 lnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPhe... 191
          ||:::::   ||||||||||||||   |||:::::|||:::|||
      723 AGAACATGGATGGCAAGGAATTGTGTAAAATGAACAAGGAGGACTTCCTC 772

      192 ArgGlnArgSerProLysGlyGlyAspValLeuHisAlaHisLeu..... 206
          |||      |||      :::   :::||||||   :::||||||
      773 CGAGCCACCTCCGCCTACAACACAGAAGTGCTGTTGTCGCCACCTCAGTTA 822

      206 ..................................................... 206

      823 CCTCAGGGAAAGTTCACTGCTGGCCTATAACACAACCTCCCATACAGACC 872

      207 ..................................Asp1eTrpLys 210
                                           |||      :::
      873 AGTCCTCACGACTGAATGTCAAGGAAGACCCTTCTTATGACTCTGTCAGG 922

      211 SerAlaAlaTrp..................................... 214
          :::||||||
      923 AGAGGAGCATGGAACAATAATATGAACTCTGGCCTCAACAAAAGTCCTCT 972

      215 ............................MetLysGluArgThrSerP 221
                                      ::::::|||   :::|
      973 CCTTGGAGGATCACAGACCATGGGCAAGAACACTGACGCAGCAGCCCAGC 1022
```

```
     221 roGlyAlaIleHisTyrCysAlaSerThrSerGluGlnSerIrpHisArg 257
         ||      :::     :::    ||||||
    1023 CAGATCCTTATCAGATCCTGGGGCCAACCAGC..................... 1054

     248 SerGlnValAspSerSerCysSerGlyGlnProIleHisLeuIrpGlnPh 284
         |||:::::   :::     |||||||||   |||:::|||||||||||
    1055 AGCCGCCTAGCAAACCCTGGGAGTGGGCAG   ATCCAGTTGTGGCAGTT 1101

     284 eLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgP 271
         !|||  .||   |||||||:::     :::   :::      |||   |
    1102 TCTCCTGGAA   CTACTGTCCGACAGCGCCAACGCCAGCTGTATCACCT 1148

     271 rpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVal 297
         ||   :::     :::|||   |||||||:::   |||      :::|||
    1149 GG   GAGGGGACCAACGGGGAGTTCAAAATGACGGACCCTGATGAGGTC 1195

     288 AlaArgLeuIrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLy 324
         ||||||   ||||||   |||||||:::::|||   |||||||||||||
    1196 GCCAGGCGCTGGGAGAGCGCAAGAGCAAGCCCAACATGAATTATGACAA 1245

     304 sLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysP 321
         ||||||||||:::::|||   ||||||   |||:::|||:::   |||
    1246 GCTGAGCGGGGCCCTCCGATACTACTATGACAAAAACATTATGACCAAAG 1295

     321 roAspIleSerGlnArgLeuVallyrClnPhe 331
                 :::::::|||   :::|||:::|||
    1296 TGCAT...GGCAAAAGGTATGCCTACAAGTTT 1324

seq_name: qb_in3:DMELGG

seq_documentation_block:
LOCUS       DMELGG       3006 bp    DNA             INV       15-JAN-1993
DEFINITION  D.melanogaster Elg gene.
ACCESSION   X68259
VERSION     X68259.1  GI:7942
KEYWORDS    DNA binding protein; elg gene; transcription factor.
SOURCE      fruit fly
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3006)
  AUTHORS   Theo,S.M., Xie,X., Smyth,F., Papas,T.S., Watson,D.K. and Schulz,R.A.
  TITLE     Molecular characterization and structural organization of D-elq, an
            ets proto-oncogene-related gene of Drosophila
  JOURNAL   Oncogene 7 (12), 2471-2478 (1992)
  MEDLINE   93096481
REFERENCE   2  (bases 1 to 3006)
  AUTHORS   Watson,D.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27 AUG-1992) D.K. Watson, NCI-FCRDC, Bldg 469, Fort
            Detrick, Frederick MD 21702, USA
FEATURES             Location/Qualifiers
     source          1..3006
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="97D"
     gene            join(568..631,709..823,885..954,1010..1724,1784..2214)
                     /gene="elq"
     exon            <568..631
                     /gene="elq"
                     /number=1
     CDS             join(568..631,709..823,885..954,1010..1724,1784..2214)
                     /gene="elq"
                     /codon_start=1
                     /protein_id="CAA48327.1"
                     /db_xref="GI:7943"
                     /db_xref="FLYBASE:FBgn0004510"
                     /db_xref="SWISS-PROT:Q04688"
                     /translation="MNSSNDLSDELIRRLSVGGALEEVIASEMFEDSIDVETEAEPMD
                     IIIVHMDIREPLSMIKSIVFQKIGVCINYYIEWIQDAQFIFSHRNTVIQCVFCFGIVQ
                     INVQIQIIKKRINIADVIKPIFAAIAALAEEVVQQLSEDSIASQKSSSSESDIKIDLK
                     RMHKEDSDEESVIGKFVKPVLNWVIQSKEKFPDGIELKIPEAANFWTHARVTYWIWAV
                     KGLELVGGNMSWSMNGLELCAMGLLQLQPLFDLLGNIEWTHEGLSEFCNEVSVVHK
                     RAFEQRKPSQFFIMSANSCGSISLSQWIMRKSYQSVKSSDSVFSITSSMNLSN
                     YTTIGSGNNGQVQLWQFLLELLTDCEHTDVIEWVGTGGEFKLIDPDRVARLWGFKKNK
                     PAMNYEKLSRALPIYYGKMISKVSKEFEAKRFTGNLSIDGLANLCIVLSKIA
                     PERLAAIEIITEDL"
     intron          632..708
                     /number=1
     exon            709..823
                     /gene="elq"
                     /number=2
     intron          824..884
                     /number=2
     exon            885..954
                     /gene="elq"
                     /number=3
     intron          955..1009
                     /number=3
     exon            1010..1724
                     /gene="elq"
                     /number=4
     intron          1725..1783
                     /number=4
     exon            1784..2552
                     /number=5
BASE COUNT      848 a    687 c    712 g    759 t
ORIGIN


alignment_scores:
            Quality:  272.50              Length:    336
              Ratio:  1.541                 Gaps:     14
 Percent Similarity:  52.976   Percent Identity: 27.381

alignment_block:
 US-09-126-945B-2 x DMELGG   ..

 Align seq 1/1  to: DMELGG  from: 1  to: 3006

      64 AspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAl 80
         |||:::|||   |||   :::::::::   ||||||   |||      ::
    1112 GACGTCCTGAAGCCCACCGAGGCCGCACTGGCTGCCTTGGCCGAGGAAGT 1161

      80 aSerSerArgGluGluProProGluGluProGluGlnCysProValIleA 97
         :   ::::::   :::||||||||||      ::::||
    1162 GGTGGGCCAGCTCTCGCCCGCCGGAAACGGCCAGCCAG.............A 1199

      97 spSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuIle 113
         |||::::::      :::   :::                     :::
    1200 AGAGTTCCTCCAGCCGAAAGCCCCATCAAAACCCCACTCAAGCGCATGCAC 1249

     114 LeuGluGluHisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
         ||:::   :::   |||:::|||:::::      |||      ||
    1250 AAGGAAGACTCCGAACAGGAATCCGTCGAGGGCAAGGATGTCAAGCCAGT 1299

     130 lLeuLys...AspIleGluThrAlaCysLys.............LeuLeuA 142
         |||:::      :::::::::      |||                 |||:
    1300 CTTGAATTGGGTCCTGGACAGCAAGTTTAAGCGAGAACAAATACGCCTGA 1349

     142 snIleIhrAlaAspProMetAspIrpSerProSerAsnValGlnLysIrp 158
         :::||                 ::::|:::   :::::::||      |||
    1350 AGATACCCCAGGCGGCCAACGAATGGACACATGCACACGTGACGTACTGG 1399

     159 LeuLeuIrpIhrGluHisGluTyrArgLeuProProMetGlyLysAlaPh 175
         |||   |||:::      |||:::::|||       ::::::   :::
    1400 CTGGAGTGGGCCGTCAAGCAGTTCGAACTAGTTGCCATCAACATGAGCGA 1449

     175 eGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheA 192
         :::::   |||:::|||||||||||||||:::::|||:::|||:
    1450 CTGGCAAATGAATGGCCAGGAACTCTGCGCAATGACGCACGAGGAATTCA 1499

     192 rgGlnArgSerPr-Leu-Gly...GlyAspValLeuHisAlaHisLeuAsp 207
         ::|||:::   |||         |||:::::::::   :::|||||||:::
```

```
     1500 ATCAGAAGCTGCCCCGTGATCCAGGCAACATCTTCTGGACCCACCTCCAG 1549

      208 IleTrpLysSerAlaAlaTrpMet.....LysGl 217
          :::   |||::::::   ::::::          :::||
     1550 CTGCTAAAAGAATGTAATTTCGTGTCCGTAGTGCACAAACGGGCGGAGGA 1599

      217 aArgThrSerPro..... 221
          |:::   :::|||
     1600 GCAGCGTAAGCCCAAGCAACCAAGAATCATGTCGGCCAATTCCATATCGA 1649

      222 .....GlyAlaIle..... 224
                |||::::::
     1650 CCAATTCAGGGGCAGCCTTTCGCTTGAACAGCGGATTATGCGGAAATCC 1699

      224 ..... 224

     1700 TATCAGTCAGTCAAGAGCTCCGATTGTAAGTAGACGTGAATGTTACAGCA 1749

      225 .....HisTyr..CysAlaSerThrSerGluGluSerTrpTh 236
                ::|||     :::   |||:::  ||||||  ||
     1750 AAGCATGAGCTAACAATATCTACTTCGTTCACAGCCGTTGAAAGTACTAC 1799

      236 rAspSer.....GluValAspSerSerCysSerG 246
          |:::|||          :::  |||:::  :::|
     1800 GTCATCCATGAATCCATCAAACTACACTACCATCGGTTCCCGGAATAATG 1849

      246 lyGlnProIleHisLeuTrpGlnPheLeuLysGlnLeuLeuLeuLysPro 262
          |||||  :::::|||||||||||||||  |||  :::|||
     1850 GACAA...GTGCAGCTCTGGCAGTTCCTACTCGAA...ATACTCACCGAC 1893

      263 HisSerTyrGlyArgPheIleArgTrpLeuAsnLysGlnGlyGlnIL 279
          :::::          |||:::|||:::::  |||  |||  ||
     1894 TGCGAGCACACGGACGTTATTGAGTGGGTGGGCACCGAG...GGAGAATT 1940

      279 eLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysA 296
          |||:::  |||      :::|||||||||||||||||  :::||||
     1941 CAAGCTCACTGACCCGCATCGAGTGGCACGCCTTTGGGGCGAGAAGAAAA 1990

      296 snArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyr 312
          ||:::|||||||||||||||:::|||||||||||||:::|||  |||
     1991 ACAAACCCGCCATGAACTACGAAAAGCTATCAAGGGCTCTCCGCTACTAT 2040

      313 TyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTy 329
          |||         :::|||  |||  :::  ::::::|||:::::||
     2041 TACGATGGGGATATGATCTCCAAGGTGTCC...GGCAAGCGATTCGCATA 2087

      329 rGlnPhe 341
          |:::|||
     2088 CAAATTT 2094

seq name: gb ov:CHKCETSZ1A

seq documentation block:
LOCUS       CHKCETSZ1A    1531 bp    mRNA            VRT       28 APR 1993
DEFINITION  Chicken c-ets-1 oncogene mRNA, complete cds.
ACCESSION   M21823
VERSION     M21823.1  GI:211466
KEYWORDS    c-ets proto oncogene; c-myc proto oncogene.
SOURCE      Chicken (strain SPAFAS), cDNA to mRNA, clones p68-[1,2].
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1531)
  AUTHORS   Leprince,D., Duterque Coquillaud,M., Li,R. P., Henry,C.,
            Flourens,A., Debuire,B. and Stehelin,D.
  TITLE     Alternative splicing within the chicken c-ets-1 locus: implications
            for transduction within the E26 retrovirus of the c-ets proto
            oncogene
  JOURNAL   J. Virol. 62, 3233-3241 (1988)
  MEDLINE   88400873
COMMENT     Draft entry and computer readable sequence [1] kindly submitted by
            D.Leprince, 16 FEB 1988.

FEATURES             Location/Qualifiers
     source          1..1531
                     /organism="Gallus gallus"
                     /db_xref "taxon:9031"
     CDS             70..1527
                     /note="c-ets-1 protein"
                     /codon_start=1
                     /protein_id "AAA48669.1"
                     /db_xref "GI:211466"
                     /translation="MMSYYMDTTIGSTGPYPLARPGVMQGASSCCEDPWMPCRLQSAC
                     CPPRSCCPPWDEAATQEVPTGLEHYSTDMECADVPLLTPSSKEMMSQALKATESGFAK
                     EQQRLGTPKDPQQWTETHVRDWVMWAVNEFSLKGVDPQKFCMNGAALCALGKECFLEL
                     APDFVGDILWEHLELLQKEEAKPYPANGVNAAYPESRYTSDYFTSYGTEHAQCVPPSE
                     FSEPSFITESYQTLHPTSSEELLSLKYENDYPSVTLRDPVQTDSLQTDYFTTKQEVVT
                     PDNMCMGKASKGKLGGQDSFESTESYDSCTDRLTQSWSSQSSFQSTQKVPSYDSFDSED
                     YPAALPNHERPKGTFEDYVRDRADMNKDKFVTPAAALAGYTGSGPTQLWQFLLELLTDK
                     SCQSFISWTGEGWEFKLSDPDEVARRWGKRKNKPKMNYEKLSRGLRYYYDKNIIHKTA
                     GKRYVYRFVCDLQSLLGYTPEELHAMLDVKPDADE"
BASE COUNT       466 a    436 c    423 g    306 t
ORIGIN      155 bp upstream of PvuII site.


alignment_scores:
            Quality:  272.00              Length:    463
              Ratio:  1.495                 Gaps:     18
 Percent Similarity:  39.309   Percent Identity: 24.190

alignment_block:
 US-09-126-945B-2 x CHKCETSZ1A   ..

 Align seq 1/1  to: CHKCETSZ1A  from: 1  to: 1531

       65 MetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaPro..... 78
          |||  |||      |||::::::  ::::::  :::|||
       73 ATCACTTACTACATGGACACAACCATTGGCAGCACGGCTCCTTATACCTTT 122

       79 .....GlyAlaSerSerArgGluGluProProG 88
                  ||||||||||||      |||  |||
      123 GGCTCGCCCTGGAGTGATGCAAGGTGCTAGCAGCTGCTGTGAGGACCCCT 172

       88 luGluPro.....GluGlnCys 93
            |||                :::  |||
      173 GCATGCCATCCACGCTGCACTCTGCCTGCTGCCCCGCCCACGCTGCTGTTGC 222

       94 ProValIleAspSerGlnAla.....ProAlaGly..... 104
          |||      ||:::  |||          |||:|||
      223 CCGCCGTGGGATGAGGCGGCCATCCAGGAAGTTCCCACTGGCCTGGAGCA 272

      104 ....SerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
              |||  ||:::      :::  :::  |||         |||
      273 CTACAGCACAGATATCGAATGTGCAGATCTGCCTTTCTTAACCCCCAGCA 322

      119 euGluGlnValGlnSerMetValValGlyGluValLeuLysAspIleGlu 135
          ::::::::::::||:::::::::     :::::::::  :::
      323 GCAAGGAAAIGATGTCTCAGGCACTGAAAGCCACCTTCAGTGGCTTCGCA 372

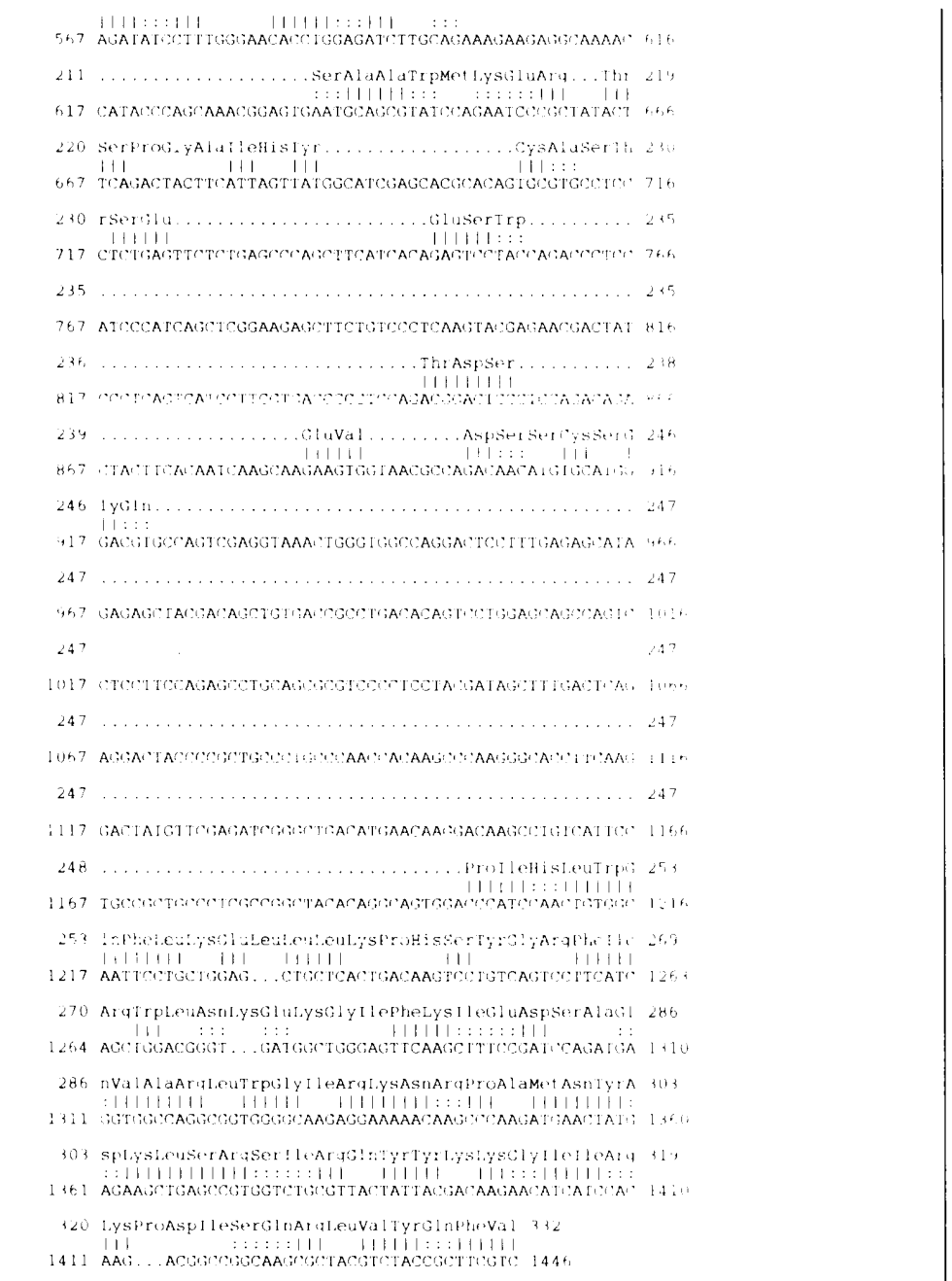
      136 ThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerPr 152
                :::  |||:::|||      ||||||::::::||||
      373 AAGGAGCAGCAGCGGCTGGGAATCCCCAAAGATCCCCAGCAGTGGACAGA 422

      152 oSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuP 169
          ::::::|||:::  |||::::::|||:::  :::::::::  |||
      423 GACGCAGTGCCGGACTGGGTGATGTGGGCAGTGAATGAGTTCAGCCTG. 471

      169 roProMetGlyLysAlaPheGlnGlu.....LeuAlaGlyLysGluLeu 183
                |||     ||||||:::      :::  |||      |||
      472 .....AAGGGAGTGGATTTCCAGAAGTTCTGCATGAACGGAGCTGCCCTG 516

      184 CysAlaMetSerGluGluGlnPheArgGlnArgSerPro...LeuGlyGl 199
          ||||||:::::::::|||  |||  :::  :::|||  :::  ||
      517 TGCGCCCTGGGCAAGGAGTGCTTCCTGGAGCTAGCGCCTGACTTTGTGGG 566

      199 yAspValLeuHisAlaHisLeuAspIleTrpLys..... 210
```



seq_name: qb_rol:MMTEL

```
seq_documentation_block:
LOCUS       MMTEL        3512 bp    mRNA            ROD       10-DEC-1996
DEFINITION  M.musculus mRNA for TEL protein.
ACCESSION   Y07915
VERSION     Y07915.1  GI:1729776
KEYWORDS    TEL protein.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3512)
  AUTHORS   Bernard,O.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3512)
  AUTHORS   Bernard,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1996) O. Bernard, 9401 Inserm, Inserm, 27 Rue J.
            Dodu, Paris, 75010, FRANCE
FEATURES             Location/Qualifiers
     source          1..3512
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /chromosome "6G2"
                     /germline
                     /cell_type="spleen"
                     /dev_stage="adult"
                     /lab_host "E.coli"
                     /clone "4"
     CDS             1215..2672
                     /codon_start 1
                     /product "TEL protein"
                     /protein_id "CAA69220.1"
                     /db_xref "GI:1729777"
                     /db_xref "SWISS-PROT:P97460"
                     /translation "MSETPAQSSIKQFRTSYIPDFESPVASHRSSIPLHVHIVEKALRM
                     EEDSTHLPTHLRIQPTYWSRDDVAQWLKWAENEFSLRPTESNKEEMNGKALLELTKED
                     FKYESPHSGDVLYEIIQQHELLNQRKSEMLESPEEPDSSIRIKEEVLLQLNHQEDNCVQ
                     RTPRTPAESVRHNPPTIELLHRPRSPITTNHRPSPDPEQQRPQRSFLDNMSRRLSPAVE
                     KAQGPRLQQENNHQETYPLSVSPVENNHCLPSSPWQESIRVIQLMPSFIMHPTLLNPR
                     HSHSVDFKQSKHSEDGMNREGKPINLSHKEDLAYLNHIMVSMSPFEEHAMPTGKIADC
                     KLLWPVVYQELSDSKYENFIPWEDKESKIFPIVDPNGIARIWGNHKNKFNMIYEEMSR
                     ALRHYYKLNIIRKEPQQRLLEFFMKTPDEIMSGPTDPLEHLESQVLDEQIYQEIEPTI
                     ASPVGWPRGNLPTGIAGGVMEAGFLGVAVKEETRE"
BASE COUNT      828 a    942 c    850 g    892 t
ORIGIN


alignment_scores:
           Quality:  270.50              Length:    531
             Ratio:  1.333                 Gaps:     20
 Percent Similarity:  38.230   Percent Identity: 22.411

alignment_block:
 US-09-126-945B-2 x MMTEL  ..

 Align seq 1/1  to: MMTEL  from: 1  to: 3512

       5 SerProGlyLeuSerSerValSerProSerHisLeuLeuLeuProProAs 21
         |||   :::|||      :::|||||||  |||         |||   ::
    1006 TCTTCAAGTCTAATGGTGCTGAGCCCGCCTCAT.....CCACAATC 1046

      21 pThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyL 68
         :::::::  |||:::          |||||:::|||
    1047 CTCGTTGACAAGAGCCCCCCCTCCCCTCGGGCGGCCGCCGCCAGCGATG 1096

      68 euGluArgArgAspTrpSer..ProSerProProAlaThrProGlu...G 53
           | ||::     :::  |||   |||   |||   |||:::   :
    1097 CTGCAAGAAACTTCTTAAACGACCGCATCCCGCTGCCCCGCCGAGCGTTT 1146

      53 lnGlyLeu.............................SerAlaPhe 58
         ::|||||                             ::::::
    1147 CTGGGTTGGGGAGAGGAAAGGAAAGTGGAAAAAAACTGAGAACTTCTGA 1196
```

```
  59 TyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAl  75
     ||||||      ||||||                                ::
1197 TCCCTTTCGCTGTGAGACATG.........................TC 1219

  75 aLysAlaProGlyAlaSerSerArgGluGlu.................P  86
     ::::::{|||:::   ||||||   ::::::                 !
1220 TGAGACTCCTGCTCAGTCTAGCATTAAGCAGGAACGAATTTCATACACGC 1269

  86 roProGluGluProGluGlnCysProValIleAspSerGlnAlaProAla 102
     |||||||||:::|||              :::   |||:::      :::
1270 CCCCAGAGAGTCCA.............GTGGCAAGCCACCGTTCCTCG 1304

 103 GlySerLeuAspLeu......ValProGlyGlyLeuThrLeuGluGluHi 117
         |||   :::      ||||||   :::|||   :::||||||
1305 ACTCCGCTTCATGTTCACACAGTGCCTCGAGCGCTCAGGATGGAGGAAGA 1354

 117 sSerLeuGluGlnValGloSerMetValValGlyGluValLeuLysAspI 134
       |||:::::::                                      :
1355 CTCGATCCAC.....................................C 1365

 134 leGlnThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrp 150
     ::   |||      :::|||                 :::|||:::   |||
1366 TGCCAACACACCTGCGTTTG.............CAGCCGATTTACTGG 1400

 151 SerProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrAr 167
     |||   ::::::|||   :::||||||   |||:::|||:::::::::
1401 AGCAGAGATGACGTAGCCCAGTGGCTCAAGTGGGCAGAAAATGAGTTTTC 1450

 167 qLeuProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuC 184
      |||   |||:::   :::          |||:::   ||||||   |||
1451 ITTAAGGCCCATTGAGAGCAACAAGTTCGAAATGAATGGCAAGGCCCTCC 1500

 184 ysAlaMetSerGluGluGlnPheArgGluArgSerProLeuGlyGlyAsp 200
         :::::::::|||:::{||||||   |||||||||   :::||||||
1501 TGCTGCTGACCAAAGAGGATTTCCGCTACCGATCTCCTCATTCAGGCGAC 1550

 201 ValLeuHis.........AlaHisLeuAspIleTrpLysSerAlaAlaTr 214
     |||||||::          ||||:
1551 GTGCTCTATGAACTCCTTCAGCATATC..................... 1577

 214 pMetLysGluArgThrSer.........................ProG 222
     :::|||:::|||   |||                          ||||
1578 CTGAAGCAGAGGAAATCTCGAATGCTCTTCTCACCATTCTTCCCCCCTG 1626

 222 ly.............................................. 222
     ||
1627 GGGACTCTATCCACACCAAGCCAGAGGTCCTCCTGCATCAGAACCATGAC 1676

 223 .....................................AlaIleHi 225
                                          :::::::||
1677 GAAGATAACTGTGTCCAGAGGACACCCAGGACGCCCGCGGAGAGCGTGCA 1726

 225 sTyrCysAlaSerThrSerGlu..................... 242
     ::::          |||   |||
1727 CCACAACCCTCCCACCATCGAACTCTTACATCGCCCTAGTCACCCATCA 1776

 242 ..............................................  242

1777 CCACAAACCACAGGCCTTCTCCTGACCCCGAACAGCAGCGGCCCCAGCGG 1826

 242 ..............................................  242

1827 TCCCCCCTAGACAACATGAGCCGCCGCCTCTCGCCAGTGGAGAAAGCCCA 1876

 242 ..............................................  242

1877 GGGGCCCAGGCTACAGCAGGAGAACAACCACCAGGAAACGTACCCCCTGT 1926

 243 .....................................GluSerTrpThr 246
                                          :::   |||
1927 CAGTGTCTCCTGTCGAGAATAATCACTGCCTGCCCTCAAGCCCCTGGCAG 1976

 237 AspSer..................................................  248
     :::|||
1977 GAGAGCACTCGAGTGATCCAGCTGATGCCCAGCCCCATCATGCACCCTTT 2026

 239 .........................GluValAspSerSerCysSer....  245
                              :::||||||   :::   |||
2027 GATCCTGAACCCCCGGCACTCGCACTCGGTGGACTTCAAACAGTCCCGGC 2076

 246 .........................GlyGlnProIleHis.........  250
                              |||:::|||||||:::
2077 ACTCCGAGGATGGGATGAATCGGGAAGGGAAGCCCATCAACCTGTCTCAT 2126

 250 ..............................................  250

2127 CGGGAGGACCTGGCTTACTTGAACCACATCATGGTCTCTATGTCCCCACC 2176

 251 ..................................................LeuT  252
                                                       ||||
2177 GGAAGAGCACGCCATGCCCATTGGCAGAATAGCAGACTGTAGACTGCTTT 2226

 252 rpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPhe  268
     ||:::::::::   :::   |||||||:::          |||   :::|||
2227 GGGATTATGTCTATCAG...TTGCTGTCTGACAGCCGGTACCAAAACTTC 2273

 269 IleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAl  285
     |||||||||   :::|||||||:::   |||||||:::|||   |||
2274 ATCCGATGGGAGGACAAAGAATCCAAAATATTCCGGATAGTGGATCCCAA 2323

 285 aGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnT  302
         :::|||||||||||||||   :::||||||||:       |||:::|
2324 CGGACTGGCTCGACTCTGGGGAAACCATAAGAACAGAACAAACATGACCT 2373

 302 yrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIle  318
     ||:::|||:::|||||:::::::|||:::|||||||||   :::|||||||
2374 ATGAGAAAATGTCCAGAGCCCTGCGCCACTACTACAAACTAAACATTATC 2423

 319 ArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal  332
     ||||||   :::   :::|||||||||:::::::::|||:::
2424 AGGAAG...GAGCCCGGACAAAGGCTTTGTTCAGGTTCATG 2462
```

```
seq_name: qb_pr5:AF116510

seq_documentation_block:
LOCUS       AF116510     1415 bp    mRNA            PRI       19 MAR 2001
DEFINITION  Homo sapiens Ets transcription factor TEL-2c (TEL 2) mRNA, complete
            cds.
ACCESSION   AF116510
VERSION     AF116510.1  GI:5442099
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1415)
  AUTHORS   Gu,X., Shin,B.H., Akbarali,Y., Weiss,A., Boltax,J., Oettgen,P. and
            Libermann,T.A.
  TITLE     Tel-2 is a novel transcriptional repressor related to the ets
            factor tel/etv-6
  JOURNAL   J. Biol. Chem. 276 (12), 9421-9436 (2001)
  MEDLINE   21153768
REFERENCE   2  (bases 1 to 1415)
  AUTHORS   Shin,B.H., Boltax,J., Gu,X., Akbarali,Y. and Libermann,T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24 DEC 1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1. .1415
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "6"
                     /map "6p21"
                     /tissue_type "pancreas"
     gene            1. .1415
```

```
                     /gene "IEL-2"
     CDS             184..965
                     /gene "IEL-2"
                     /codon_start=1
                     /product="Ets transcription factor IEL 2e"
                     /protein_id="AAD43252.1"
                     /db_xref="GI:5442100"
                     /translation="MNSPALCILIFFLPPPAPSSPVYFSLYFPLPSALVPSP
                     PGGIFRLKTPIQHSPVPPFEVIGPSQMDIRRGHILIQPFPPGLISNFGHLDDPGLARWI
                     PGKEESLNLCHCALLGCRTQGVCSFFAMPQAFIDGRIAYTRLLWDTVTCLLLVTSFG
                     YIKWEDKDAKIFFVVDPNGLARLWGNHKNSVNMTYEKMSPALIHYYFLNIIKEEPCQF
                     LLFRFLKTPGKMVQDKHSHLEPLFSQEQDRIEFKDKRPEISP"
BASE COUNT       438 a    418 c    415 g    244 t
ORIGIN


alignment_scores:
            Quality:  269.50              Length:    263
              Ratio:  2.105                 Gaps:      5
 Percent Similarity:  48.669   Percent Identity: 25.856

alignment_block:
 US-09-126-945B-2 x AF116510  ..

 Align seq 1/1  to: AF116510  from: 1  to: 1415

     142 AsnIleThrAlaAspProMetAspTrpSerProSerAsnValGlnLysIr 188
         |||          ::|||       ||||||   :::::|||         ||
      75 AATGCAGGCATCCAGCCCGCACTCTGGAGCAGGCAGGACGTGCTGCACTG 124

     158 pLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaP 178
         ||||  |||:::|||:::::::|||   ||||||      ::::::
     125 GCTGCGCTGGGCAGAGCAGGAGTACTCTCTGCCATGCACCGCGGCAGCACG 174

     175 beGlnGlnLeuAlaGlyLysGlnLeuCysAlaMetSerGlnGlnGlnPhe 191
             ||:::   |||:::   ||||||   :::::::::::::::::|
     175 GGTTCGAGATGAACGGACGCGCCCTCTGCATCCTCACCAAGGACGACTTC 224

     192 ArgGlnArgSerProLeuGlyGlyAspValLeuHis..............  203
         |||:::|||:::|||   :::|||||||||||:::
     225 CGGCACCGTGCGCCCAGCTCAGGTGACGTCCTCTATGAGCTGCTCCAGTA 274

     203 ..................................................  203

     275 CATCAAGACCCAGCGGCGAGCCCTGGTGTCTGGGCCCTTTTTTGCAGGCA 324

     203 ..................................................  203

     325 TCTTCAGGCTGAAGACGCCCACCCAGCACTCTCCAGTCCCCCCCGGAAGAG 374

     203 ..................................................  203

     375 GTGACTGGCCCCTCTCAGATGGACACCCGAAGGGCCACCTGCTGCAGCC 424

     204 ..........................AlaHisLeuAspIleTrpLysS 211
                                  :::|||||||||
     425 ACCAGACCCAGGGCTTACCAGCAACTTCGGCCACCTGGATGACCCTGGCC 474

     211 erAlaAlaTrpMetLysGluArgThrSerProGlyAlaIleHisTyrCys 227
           |||   |||     :::   :::         :::    ::::||
     475 TGGCAAGGTGGACCCCTGGCAAGGAGGAGTCCCTCAACTTATGTCACTGT 524

     228 AlaSer.......................ThrSerGlnGluSerTrpTh 249
         ||::::                      |||       :::
     525 GCAGAGCTCGGCTGCACCACCCACGGGCTCTCTTCCTTCCCCCCATGGC 574

     250 rAspSerGlnValAspSerSerCysSerGlyGlnProLeuHisLeuTrpI 254
          ::::::   :::|||:::      :::          :::   ||||||:
     575 GCAGGCCCCCATTGACGGCAGGATCGGTGACTGCCGCCTG...CTGTGGA 621

     253 lnPheLeuLysGluLeuLeuLysProHisSerTyrGlyArgPheIle 269
         ::::::::   :::||||||||||      :::   |||      ::::||
     622 ATTACGTGTATCAGCTGCTCCTTGATACCCGA...TATGAGCGCTACATT 668

     270 ArgTrpLeuAsnLysGlnLysGlyIlePheLysIleGluAspSerAlaCl 286
         :::|||   :::|||:::       ||||||||||:::   |||
     669 AAGTGGGAAGACAAGGACGCCAAGATCTTCCGAGTTGTGGATCCAAATGG 718

     286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 303
         :::|||||||||||||||   :::||||||||      ||||:|||:
     719 GTTCGCCAGACTCTGGGCAAAGCACAACAACCCCCTCAACATGAACTACG 768

     303 spLysLeuSerArgSerIleArgInTyrTyrLysLysGlyIleIleArg 319
         ::|||:::|||||:::::|||:::|||||||||   :::||||||:::
     769 AGAAGATGTCTCGTGCCCTGCGCCACTATTATAAGCTTAATATCATTAAG 818

     320 LysProAspIleSerGlnArgLeuValTyrGlnPheVal 352
         |||   :::   :::|||:::|||:::::::::|||:::
     819 AAG...GAACCGGGGCAGAAACTCCTGTTCAGATTTCTA 854

seq_name: gb_pr5:AF218365

Seq_documentation_block:
LOCUS       AF218365     1400 bp    mRNA            PRI       19-MAR-2001
DEFINITION  Homo sapiens Ets transcription factor IEL2E (IEL2) mRNA, complete
            cds.
ACCESSION   AF218365
VERSION     AF218365.1  GI:7274371
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1400)
  AUTHORS   Gu,X., Shin,B.H., Akbarali,Y., Weiss,A., Boltax,J., Oettgen,P. and
            Libermann,T.A.
  TITLE     Tel-2 is a novel transcriptional repressor related to the ets
            factor tel/etv-6
  JOURNAL   J. Biol. Chem. 276 (12), 9421-9436 (2001)
  MEDLINE   21153768
REFERENCE   2  (bases 1 to 1400)
  AUTHORS   Shin,B.-H., Gu,X., Akbarali,Y., Weiss,A., Boltax,J., Oettgen,P. and
            Libermann,T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1400
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21"
                     /tissue_type="prostate gland"
     gene            1..1400
                     /gene="IEL2"
     CDS             26..979
                     /gene="IEL2"
                     /function "transcriptional repressor"
                     /note="related to the Ets factor TEL/ETV-6"
                     /codon_start=1
                     /product "Ets transcription factor IEL2E"
                     /protein_id="AAF44742.1"
                     /db_xref "GI:7274372"
                     /translation="MQFGELAISPTSPVAAMPTPGLDVQAKCPAQLNGLGGGLCKLP
                     GRLRTQPALWSKEDVLHWLRWAFQEYSLPCTAEHGFEMNGKALCTLTKDDFRHDAFSS
                     CDVLYELLQYIKTPQKRALVCGPPSCTPFLKTPIQHSPVPPFEVIGPSQMDIRKGHLI
                     QPFPPGLISNFGHLDDPGLARWTIPGKEESLNLCHCAELGCRTQGVCSFFAMPQAFIDG
                     RIAYCRLLWDTVYVQLLLLVTSFEDYIKWEDKDAKIFFVVDPNGLARLWGNHKNSVNMTY
                     EKMSPALIHYYFLNIIKEEPCQFLLFRFLKTPGKMVQDKHSHLEPLFSQEQDRIEFKDK
BASE COUNT       290 a    271 c    344 g    296 t
ORIGIN


alignment_scores:
            Quality:  268.50              Length:    318
              Ratio:  1.755                 Gaps:      8
```

```
 Percent Similarity:  48.114   Percent Identity: 24.528

Alignment block:
 US-09-126-945B-2 x AF218865   ..

 Align seq 1/1  to: AF218865  from: 1  to: 1400

      86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
         ||||||          |||          :::|||:::      :::||||::
      17 CCACCTGTAATGCAGGAGGGAGAATTGGCTATTTCTCCTATAAGCCCTGT 66

     102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
         ::::::::::          |||   |||       ::::::   :::
      67 GGCAGCCATGCCT......CCCCTAGGCACCCACGTGCAAGCCAGATGTG 110

     119 euGluGlnValGlnSerMetValValGlyGluValLeuLysAspIleGlu 135
             |||:::::::   :::       |||   :::
     111 AAGCTCAAATTAACCTGCTGGGTGAAGGGGGGATC.............. 145

     136 ThrAlaCysLysLeuLeu...AsnIleThrAlaAspProMetAspTrpSe 151
             |||||||||       ::::::      :::|||      |||||
     146 ......TGCAAGCTGCCAGGAAGACTCCGCATCCAGCCCGCACTGTGGAG 189

     151 rProSerAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 168
         |   ::::::|||      ||||||   ||:::|||||::::|||   |
     190 CAGGGAGGACGTGCTGCACTGGCTGCGCTGGGCAGAGCAGGAGTACTCTC 239

     168 euProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCys 184
         |||||       ::::::      |||:::   ||::::   ||||||
     240 TGCCATGCACCGCGGAGCACGGGTTGGAGATGAACGGACGCGCCCTCTGC 289

     185 AlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspVa 201
         ::::::::::::::::|||||:::|||:::|||   :::|||||||||
     290 ATCCTCACCAAGGACGACTTCCGGCACCGTGCGCCCAGCTCAGGTGACGT 339

     201 lLeuHis..................................................... 203
         ||||:::
     340 CCTGTATGAGCTGCTCCAGTACATCAAGACCCAGCGGCGAGCCCTGGTGT 389

     203 ..... 203

     390 GTGGGCCCTTTTTTGGAGGGATCTTCAGGCTGAAGACGCCCACCCAGCAC 439

     203 ..... 203

     440 TCTCCAGTCCCCCCGGAAGAGGTGACTGGCCCCTCTCAGATGGACACCCG 489

     204 ..... A 204
                                                         :
     490 AAGGGGCCACCTGCTGCAGCCACCAGACCCAGGACTTACCAGCAACTTCG 539

     204 laHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThr>> 227
         ::|||||||||          |||   |||       :::   :::
     540 GCCACCTGGATGACCCTGCCCTGGCAAGGTGGACCCCTGGCAAGGAGGAG 589

     221 ProGlyAlaIleHisTyrCysAlaSer..................... 229
                 :::   :::|||||:::
     590 TCCCTCAACTTATGTCACTGTGCAGAGCTCGGCTGCAGGACCCAGGGGCT 639

     230 .ThrSerGluGluSerTrpThrAspSerGluValAspSerSerCysSerG 246
          |||      :::      ::::::   :::|||:::      :::
     640 CTGTTCCTTCCCCGCGATGCCGCAGGCCCCCATTGACGGCAGGATCGCTG 689

     246 lyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysPro 262
             :::   ||||||:::::::::   :::||||||||||
     690 ACTGCCGCCTG...CTGTGGGATTACGTGTATCAGCTGCTCCTTGATACC 736

     263 HisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePh 279
     :::   |||      :::|||:::|||   :::|||:::      |||||
     737 CGA...TATGAGCCCTACATCAAGTGGGAAGACAAGGACGCCAAGATCTT 783

     279 eLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysA 299
         |:::::   |||         :::|||||||||||||||   :::||||
```

```
     784 CCGAGTTGTGGATCCAAATGGGCTCGCCAGACTCTGGGGAAATCACAAGA 833

     296 snArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyr 412
         |||||       |||:::|||:::|||:::|||||:::::::|||:::|||
     834 ACCGGGTGAACATGACCTACGAGAAGATGTCTCGTGCCCTGCGCCACTAT 883

     413 TyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTy 429
         ||||||   :::|||||||:::|||   :::   :::|||:::|||||:::
     884 TATAAGCTTAATATCATTAAGAAG...GAACCGGGGCAGAAACTCCTGTT 930

     429 rGln 430
         ::::
     931 CAGA 934

seq_name: qb_rol:MUSETS2

seq_documentation_block:
LOCUS        MUSETS2      1552 bp    mRNA            ROD       12-JUN-1993
DEFINITION   Mouse erythroblastosis virus oncogene homolog 2 (ets 2) mRNA,
             complete cds.
ACCESSION    J04103
VERSION      J04103.1  GI:193193
KEYWORDS     ets 2 gene.
SOURCE       Mouse (Balb/c 3T3), cDNA to mRNA.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 1552)
  AUTHORS    Watson,D.K., McWilliams,M.J., Lapis,P., Lautenberger,J.A.,
             Schweinfest,C.W. and Papas,T.S.
  TITLE      Mammalian ets-1 and ets-2 genes encode highly conserved proteins
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 85, 7862-7866 (1988)
  MEDLINE    89042086
COMMENT      Draft entry and computer readable sequence [1] kindly submitted by
             J.Lautenberger 07 OCT 1988.
FEATURES             Location/Qualifiers
     source          1..1552
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             138..1544
                     /note="ets2 protein"
                     /codon_start=1
                     /protein_id="AAA37581.1"
                     /db_xref="GI:309221"
                     /translation="MNDFGTKNMDQVAPVANSFRGTLKRQPAFDFFDGSLFAVLPSLS
                     EDQTLQEVPTGLDSVSHDSASCTELPLEIPQSKAVMSQAEKATFSGFQKEQRRLSTFKN
                     PWLWSEQQVCQWLLWATNEFSLVNVNLHQFGMNGQMLCNLGKERFLELADPEVGGLLW
                     EHLEQMIKENQEKTEDQYEENSHLNAVFHWTNSNTLGFSMEQAPYGMQAPNYFKQNEL
                     DSMCPPSATPAAIGSELQMLPKSRLNTVNVNYCSISQDFPSSNVNLLNNNSGKPKTHHD
                     SPENGGDSFESSDSLLKSWNSQSSLLDVQRVPSFESFEEECSQSLCLSKLLMSFKDYI
                     QERSDPVEQGKPVIPAAVLAGFTGSGPIQLWQFLLELLSDKSCQSFISWIGDHGWEFKL
                     AEFFKVAEEWCERKEKPEMNYEKLSRGLEYYYERKCITPERVYRFVCDIQNLDG
                     FTPEELHAILGVQPDTED"
BASE COUNT       360 a    471 c    407 g    314 t
ORIGIN       Chromosome 16.


alignment_scores:
            Quality:  267.50              Length:    497
              Ratio:  1.268                 Gaps:     19
 Percent Similarity:  42.455   Percent Identity: 22.938

alignment_block:
 US-09-126-945B-2 x MUSETS2   ..

 Align seq 1/1  to: MUSETS2  from: 1  to: 1552

      11 ValSerProSerHisLeuGluGluProProAspThrValSerArgThrGl 27
         |||||||||   ||||||      |||                  ||
       2 GTCAGTCCCCCGCCACCTCCCCGGCCCCGCGCGCCGGATGGGCCCTACGG 51

      27 yLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAsp...T 44
         ||||      ::::::   |||   :::       :::||||||||||
```



```
 52 YCIC.....GTCTCGCCCGGCCTTGCGCGCCGGGACCGCCGCGATCTCC 95

 43 rpSerPro.SerProProAlaThrProGlu.....GlnGlyLeuSerAl 57
      |||||'    |||||||||   |||          :::::::::
 96 TCTCCCCGCCGCCCTCCGGCTGGCCCTGCCTGCTGCGGCGCGATCAATGA 145

 57 aPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpA 74
    |||   ::::::   :::|||:::::   |||      :::|||:::
146 CTTTGGAATCAAGAACATGGACCAAGTGGCCCCTGTCGCCAACAGTTTTC 195

 74 EaAlaLysAlaProGlyAlaSerSerArgGluGluProProGlu..... 88
                 ||||::   :::|||:::         :::
196 GT.....GGGACACTCAAGCGCCAGCCAGCCTTTGACACCTTC 233

 89 GluProGluGlnCysProValIleAspSerGlnAlaProAlaGlySerLe 105
    :::  :::        ||:::  |||  :::          ::::|
234 GATGGCTCTCTGTTTGCTGTGCTCCCTTCTCTCAGTGAAGATCAGACACT 283

105 aAspLeuValProGlyGlyLeuThrLeuGluGluHisSer..... 118
    ::::  ||||||  ||||||         ::|||:::
284 CCAAGAAGTGCCCACGGGCCTGGATTCTGTTCCCATGACTCGGCCAGCT 333

119 .............LeuGluGlnValGlnSerMetValValGlyGluVal 130
                |||         ::::::   ||:::::::::::::
334 GCGAGCTGCCTTGCTCACTCCCTGCAGCAAGGCAGTGATGAGCCAAGCC 383

131 LeuLys.............AspIleGluThrAlaCysLysLeuLeuAsnIl 145
    ||||||             ::::::      :::   ||:::|||
384 TTAAAAGCCACCTTCAGTGGCTTCCAAAAGGAGCAACGACGTCTTGGCAT 433

146 eThrAlaAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuL 160
    |      ::::||      ||||||   :::::|||   :::|||||
434 CCCCAAAAACCCCTGGCTGTGGAGCGAGCAGCAGGTGTGCCAGTGGCTTC 483

160 euTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGln 176
    |||||:::   :::::::::   |||       ::::::
484 TCTGGGCCACCAACGAGTTCAGCCTGGTGAATGTGAACCTGCACCAGTTT 533

177 GluLeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgSl 193
    :::   ||:::   ||:|||   :::::::::|||:::|||   ::
534 GGCATGAACGGCCAGATGCTGTGTAACCTCGGCAAGGAGGCGCTTCCTGGA 583

193 nArgSerPro...LeuGlyGlyAspValLeuHisAlaHisLeu..... 206
    :  ::|||   :::  ||||||:::|||      ||||||
584 GCTGGCGCCTGACTTGTGGGTGACATCCTCTGGGAACATCTAGAGCAGA 633

207 ..........................AspIleTrpLysSerAlaAla 218
                             |||   :::::::::   :::
634 TCATCAAAGAGAACCAAGAAAAGACAGAAGACCAATATGAGGAAAACTCT 683

219 .....................IrpMetLysGluArgThr..... 233
                         |||:::::::::::::||
684 CACCTCAACGCGGTTCCTCATTGGATCAACAGCAATACATTAGGCTTCAG 733

234 ..................................................... 234

734 CATGGAACAGGCTCCATATGCAATGCAGGCACCAAACTACCCCAAAGACA 783

235 ..............................SerProGlyAla..... 248
                                 :::|||:::|||
784 ATCTCCTGGACAGCATGTGCCCGCCATCGGCCACCCCTGCAGCTCTGGGC 833

249 .....................................IleHi 255
                                         :::::
834 TCTGAGCTCCAGATGTTGCCCAAGTCTCGGCTCAACACCGTCAATGTCAA 883

256 sTyrCysAlaSerThrSerGln..... 272
    :|||||||:   :::::::::
884 TTACTGTTCCATCAGCCAGGACTTCCCCAGCAGCAACGTGAATTTGTTCA 933

273 ..................................................... 273

934 ACAACAATTCTGGAAAACCCAAGGACCACGACTCTCCAGAGAACGGTGGC 983

      288 ..............................GluSerIrpThrAsp..... 297
                                       :::|||||:::::::
      984 GACAGCTTCGAGAGCTCCGACTCGCTGCTGAGGTCCTGGAACAGCCAGTC 1033

      298 ..................................SerGluValAspSerS 313
                                           ::::::   :::::::
     1034 GTCCCTACTGGATGTACAGCGGGTACCTTCCTTCGAGAGCTTTGAGGAGG 1083

      313 erCysSer..................................... 319
          :::|||||
     1084 ACTGTAGCCAGTCTCTGTGCCTCAGTAAGCTGACCATGTCCTTCAAGGAC 1133

      320 ..................................................... 320

     1134 TACATCCAAGAGAGGAGCGACCCAGTCGAGCAAGGCAAACCAGTTATTCC 1183

      321 ..........................GlyGlnProIleHisLeuIrpG 334
                                   :::   ||||||:::::::::::
     1184 TGCAGCAGTACTGGCTGGCTTCACTGGAAGCGGACCAATCCAGTTGTGGC 1233

      335 InPheLeuLysGluLeuLeuLysProHisSerTyrGlyArgPheIle 351
          ||||||||   |||   ||||||:::      |||         ||||||
     1234 AGTTTCTTCTGGAG...CTACTCTCTGACAAGTCCTGTCAATCTTTCATC 1280

      352 ArgIrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 368
          |||   :::  :::          ||||||:::   |||      ::
     1281 AGCTGGACGGGG...GATGGATGGGAGTTCAAGCTTGCTGACCCCGATGA 1327

      369 nValAlaArgLeuIrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 385
          :||||||||   ||:|||   ||||||||:::::|||   ||||||||:
     1328 GGTTGCCCGCCGCTGGGGCAAGAGGAAAAAATAAACCAAAGATGAACTACG 1377

      386 spLysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleIleArg 402
          ::|||||||||||:::::::||   ||||||   ||:::|||||||:::
     1378 AGAAGCTGAGCCGGGGCTTACGTTACTACTACGACAAGAACATCATCCAC 1427

      403 spLysProAspIleSerGlnArgLeuValTyrGlnPheVal 432
          |||   :::   ::::::|:|   |:|:|||::|:||||
     1428 AAGACTTCG...GGCAAGCGCTACGTGTACCGTTTCGTA 1463

seq_name: qb_pr10:HUMFLI1A

seq_documentation_block:
LOCUS       HUMFLI1A     1932 bp    mRNA            PRI       31 DEC-1994
DEFINITION  Human FLI-1 mRNA, complete cds for two alternate splicings.
ACCESSION   M93255
VERSION     M93255.1  GI:182659
KEYWORDS    FLI-1; alternative splicing; ets oncogene family; transcription
            factor.
SOURCE      Homo sapiens (tissue library: lambda gt11 HEL 21) blood cDNA to
            mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1932)
  AUTHORS   Bromas,R., May,W., Denny,C., Raskind,W., Moore,J., Maki,R.A.,
            Beck,E. and Klemsz,M.J.
  TITLE     Human FLI 1 localizes to chromosome 11q24 and has an aberrant
            transcript in neuroepithelioma
  JOURNAL   Biochim. Biophys. Acta 1172 (1-2), 155-158 (1993)
  MEDLINE   93176799
FEATURES             Location/Qualifiers
     source          1..1932
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HEL"
                     /cell_type="erythroleukemia"
                     /tissue_type="blood"
                     /tissue_lib="lambda gt11 HEL 21"
                     /map="11q24-25"
     mRNA            join(1..401,576..1932)
                     /gene="FLI-1"
```

```
                     /note "alternate"
     mRNA            1. .1932
                     /note "alternate"
     gene            join(1. .401,576. .1932)
                     /gene "FLI 1"
     CDS             join(174. .401,576. .1700)
                     /gene "FLI-1"
                     /note "alternate"
                     /codon_start 1
                     /protein_id "AAA58479.1"
                     /db_xref "GI:182660"
                     /translation "MEGGLAGERAKESPVDCSVGKCYELVPCCIENFMNYNGVMEIKN
                     GPPPPNMTTNERRVIVPADATLVTQEHVRQWLEWAIKEYSLMEIDTSFFQNMDGKELC
                     KMNKEDFLRATTLYNTEVELLSHLSYLRESSLLAYNTTSHITDQSSRLSVKEDPSYDSVR
                     RGAWGNNMNSGLNKSPPLGGAQTISKNTEQRPQPDPYQILGPTSSRLANPGSGQIQLW
                     QFLLELLSDSANASCITWEGTNGEFKMTDPDEVARRWGQRKSKPNMNYDKLSRALRYY
                     YDKNIMTKVHGKRYAYKFDFHGIAQALQPHPTESSMYKYPSDISYMPSYHA
                     HQQKVNFVPPHPSSMPVTSSSFFGAASQYWTSTGGIYPNPNVPKHPNTHVPSHLGSYY
                     "
     CDS             543. .1700
                     /gene "FLI 1"
                     /note "alternate"
                     /codon_start 1
                     /product "FLI 1"
                     /protein_id "AAA58480.1"
                     /db_xref "GI:182661"
                     /translation "MEGGLAGERAKESPVDCSVGKCYELVPCCIENFMNYNGVMEIKN
                     GPPPPNMTTNERRVIVPADATLVTQEHVRQWLEWAIKEYSLMEIDTSFFQNMDGKELC
                     KMNKEDFLRATTLYNTEVELLSHLSYLRESSLLAYNTTSHITDQSSRLSVKEDPSYDSVR
                     RGAWGNNMNSGLNKSPPLGGAQTISKNTEQRPQPDPYQILGPTSSRLANPGSGQIQLW
                     QFLLELLSDSANASCITWEGTNGEFKMTDPDEVARRWGQRKSKPNMNYDKLSRALRYY
                     YDKNIMTKVHGKRYAYKFDFHGIAQALQPHPTESSMYKYPSDISYMPSYHAHQQKVNF
                     VPPHPSSMPVTSSSFFGAASQYWTSTGGIYPNPNVPKHPNTHVPSHLGSYY"
BASE COUNT       487 a     554 c     519 g     372 t
ORIGIN      chromosome 11q23.3-25.


alignment scores:
            Quality:  267.50              Length:     395
              Ratio:  1.372                 Gaps:      19
 Percent Similarity:  49.367   Percent Identity:  26.835

alignment block:
 US-09-126-945B-2 x HUMFLI1A  ..

 Align seq 1/1  to: HUMFLI1A  from: 1  to: 1932

      19 ProProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAl  65
         ||||||   |||||||||       ||||||||||||||||||||   ||
     320 CCCCCTCCCACCACAGCA......GGAGTGGATCAATCAGCCAGTGAGGG  368

      66 a.........ValGlyLeuGluArgArgAspTrpSerProSerProProA  99
         :         :::|||||:       :::::|||       :::|||   |
     364 TCAACGTCAAGCGGGTGTATGACCACATGAATGGATCCAGGCGCCACGTG  413

      99 laThrProGluGlnGlyLeuSer.AlaPheTyrLeuSerTyrPheAspMe 145
         ||:::   :::          ||| ||||||||   :::
     414 CCTCATTAAAG.........AGCAGCCTTTTATGCTGGGCTTCAC..... 449

     146 tLeuTyrPro.....................................GluA 160
          |||||||||                                     :::
     450 .CTGTACCCACCCTGAGCTGCAGCCGTGGAGCCCCATGGCCGCACGCAGC 498

     161 spSerSerTrp.....................AlaAlaLysAlaPro 178
             |||                       ||||||||:::
     499 GCTTGCGCTGGCTGGAGGAGCCACGGTGCTTGGGAGCTGCAAGAATGGAG 548

     179 GlyAlaSerSerArgGluGluProProGluGluProGluGlnCysProVa 225
         ||:::   :::  |||:::      |||:::|||   :::|||   ||
     549 GGAGCACTGGCAGGCGAGCGGGCGAGGGAGTCTCCGGTGGACTGCAGCCT 598

     226 tIleAspSerGluAlaProAlaGlySerLeuAspLeuValProGlyGlyL 272

         |                  :::             ||||||   ||||||
     599 T.................AGCAAATGCAGCAAGCTGGTGGGCGGAGGCG  640

     273 euThr.......................LeuGluGluHisSerLeuGlu 320
         :::                       :::::|||   :::
     641 AGTCCAACCCCATGAACTACAACAGCTATATGGACGAGAAGAATGGCCCC  690

     321 GlnValGlnSerMetValValGlyGluValLeuLysAspIleGluThrAl 367
                 :::|||||||||||||
     691 CCTCCTGCCAACATGACCACCAACGAG.......................  717

     367 aCysTyrLeuCysAspIleGluAlaAspAspMetAspIleTrpSerProSerA 414
         :::   :::   :::   ||||||                 :::   ::::
     718 ....AGGAGAGTCATCGTCCCGGCAGACGCCACACTGGTGACACAGGAGC  768

     414 snValGlnLysIrpLeuLeuIrpThrGluHisGlnTyrArgLeuProPro 460
         ::|||::::::|||||   ||||::      :::|||   |||
     754 ATGTGAGGCAATGGCTGGAGTGGGCCATAAAGGAGTACAGCTTGATGGAG  803

     461 MetGlyLysAla...PheGlnGluLeuAlaGlyLysGluLeuCysAlaMe 506
         :::     :::   ||||||||||||   |||||||||||||||   ||
     804 ATGGACACATCCTTTTTCCAGAACATGGATGGCAAGGAACTGTGTAAAAT  853

     507 tSerGluGluGlnPheArgGlnArgSerProLeu...GlyGlyAspValL 552
         |:::::|||:::|||   :::   :::   |||   :::   :::||||
     854 GAACAAGGAGGACTTCCTCCGCGCCACCACCCTCTACAACACGGAAGTGC  903

     553 euHisAlaHisLeu..................................... 566
         ||   :::||||||
     904 TGTTGTCACACCTCAGTTACCTCAGGGAAAGTTCACTGCTGGCCTATAAT  953

     567 ..................................................... 606

     954 ACAACCTCCCACACCGACCAATCCTCACGATTGAGTGTCAAAGAAGACCC 1003

     607 .......AspIleIrpLysSerAlaAlaIrpMetLysGluArgIhrSerP 651
                |||     :::   :::||||||   ::::::   :::|||
    1004 TTCTTATGACTCAGTCAGAAGAGGAGCTTGGGGCAATAACATGAATTCTG 1053

     652 ro.....................GlyAlaIleHisTyrCysAlaSerThr 690
                                |||:::   :::          :::::
    1054 GCCTCAACAAAAGTCCTCCCCTTGGAGGGGCACAAACGATCAGTAAGAAT 1103

     691 SerGluGlu.............................SerIrpIhrAs 717
         :::|||:::                          :::   ||||:
    1104 ACAGAGCAACGGCCCCAGCCAGATCCGTATCAGATCCTGGGCCCGACCAG 1153

     718 pSerGluValAspSerSerCysSerGlyGlnProIleHisLeuIrpGlnP 764
         :|||::::::   :::       |||||||||   |||:::|||||||||
    1154 CAGTCGCCTAGCCAACCCTGGAAGCGGGCAG...ATCCAGCTGTGGCAAT 1200

     765 heLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArg 810
         |||||   |||   |||||||:::       :::   :::       |||
    1201 TCCTCCTGGAG...CTGCTCTCCGACAGCGCCAACGCCAGCTGTATCACC 1247

     811 TrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnVa 857
         |||   :::      :::|||   |||||||:::   |||      :::||
    1248 TGG...GAGGGGACCAACGGGGAGTTCAAAATGACGGACCCCGATGAGGT 1294

     858 lAlaArgLeuIrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspL 904
         |||||||   ||||||   |||||:::::|||   ||||||||||||||
    1295 GGCCAGGCGCTGGGGCAGCGGAAAAGCAAGCCCAACATGAATTACGACA 1344

     904 ysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleIleArgLys 950
         ||||||||||||:::::|||   ||||||   |||:::|||:::   |||
    1345 AGCTGAGCCGCAGCATCCGCGAATACTACAAGAAAAACATTATCAGCAAA 1394

     951 ProAspIleSerLnArgLeuValTyrGlnPhe 981
                :::::|||   :::|||:::|||
    1395 CTGCAC...GCAAAAGATATGCTTACAAATTT 1424

seq_name: qb_ov:AF177538
```

```
seq_documentation_block:
LOCUS       AF177538     1245 bp    mRNA            VRT       12-OCT-1999
DEFINITION  Danio rerio flil protein (flil) mRNA, partial cds.
ACCESSION   AF177538
VERSION     AF177538.1  GI:6019200
KEYWORDS    .
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 1245)
  AUTHORS   Thompson,M.A., Ransom,D.G., Pratt,S.J., MacLennan,H., Kieran,M.W.,
            Detrich,H.W., Vail,B., Huber,T.L., Paw,B., Brownlie,A.J.,
            Oates,A.C., Fritz,A., Gates,M.A., Amores,A., Bahary,N.,
            Talbot,W.S., Her,H., Beier,D.R., Postlethwait,J.H. and Zon,L.I.
  TITLE     The cloche and spadetail genes differentially affect hematopoiesis
            and vasculogenesis
  JOURNAL   Dev. Biol. (NY) 197, 248-269 (1998)
REFERENCE   2  (bases 1 to 1245)
  AUTHORS   Thompson,M.A., Ransom,D.G., Pratt,S.J., MacLennan,H., Kieran,M.W.,
            Detrich,H.W., Vail,B., Huber,T.L., Paw,B., Brownlie,A.J.,
            Oates,A.C., Fritz,A., Gates,M.A., Amores,A., Bahary,N.,
            Talbot,W.S., Her,H., Beier,D.R., Postlethwait,J.H. and Zon,L.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (13 AUG-1999) Hematology, Children's Hos Boston, 300
            Longwood Ave, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1..1245
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /chromosome="18"
     gene            <1..>1245
                     /gene="flil"
     CDS             <1..>1245
                     /gene="flil"
                     /codon_start=3
                     /product="flil protein"
                     /protein_id="AAF01966.1"
                     /db_xref="GI:6019201"
                     /translation="EDQSLEPPPYAAAAPIPKTDMTASCIQEVSQIHKINPIPPQQEW
                     INQPVRVNVKEEYDEINGSEESFVDCEVCKCNKMVGCTLASQMNYICYMHEKCAPTPN
                     MTINERRVIVPADPSLWSPDHVRQWLDWAIKEYGLQEIDTAMFHSIDGEELCKMSKDD
                     FLRLTSVYNTPVLLSHLNYLRFSSSSISYNTPSHADQSPRLAAKDDASYDAVRRTGWS
                     NNMHSGKGSPTVVSQSVSKNPDQPRPQPDPYQTLGPTSSRLANPGSGQTQLWQFLLEL
                     LSDSANAGCITWEGTNGEFKMTDPDEVARRWGERKSKPDMNYDKLSRALRYYYDKIIM
                     TKVHGKPYAYKIDFHGIAQALQGHPIESIMYKYPSELQYVFSYHADQQPVNFVSPHPP
                     SMPVISSNFFGPTTPYWSSPPR"
BASE COUNT       321 a     394 c     309 g     221 t
ORIGIN


alignment_scores:
            Quality:  266.50              Length:     228
              Ratio:  2.239                 Gaps:       8
 Percent Similarity:  52.193   Percent Identity:  42.895

alignment_block:
 US-09-126-945B-2 x AF177538   ..

 Align seq 1/1  to: AF177538  from: 1  to: 1245

     143 IleThrAlaAspProMetAspTrpSerProSerAsnValGluLysTrpLe 194
         :::   ||||||||       ||||||||||:::||||||:::||||||
     536 GTCCCCGCACACCCGTCTCTGTGGTCTCCGGATCACGTGCGGCAGTGGTT 585

     195 uLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAla...P 175
         |   ||:::       :::|||   |||       :::       |||   |
     586 AGACTGGGCCATTAAGGAGTACGGTCTTCAGGAGATCGACACGGCCATCT 635

     175 heGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGlnGluGlnPhe 191
         ||:::::       ||||||||||||||   ||||||:::::::|||
     636 TTCACAGCACAGATGGAAAAGAGCTCTGCAAGATGAGCAAAGACGACTTC 685

     192 ...ArgGlnArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu.. 206
            |||      |||      :::   :::||||||   :::||||||
     486 CTGAGACTCACCAGCGTTTATAACACCGAGGTCCTGCTCTCACATCTCAA 535

     206 ..................................................... 206

     536 TTACCTCAGGGAAAGTAGCTCATCGATATCATACAACACGCCATCTCACG 585

     207 .....................................------------Aspile 208
                                                          |||
     586 CTGACCAGTCCCCTCGCCTGGCTGCAAAAGACGATGCGTCTTATGATGCT 635

     209 TrpLysSerAlaAlaTrpMetLysGlu.............ArgPheSerPr 221
            :::   :::::::|||   ::::::              :::   |||||
     636 GTACGGAGCACGGGCTGCTCGAACAACATGCACACGCGAAAAGGCTTCTCC 685

     221 cGlyAlaIleHisTyrCysAlaSerThrSerGluGln.............. 233
         |  ::::::      :::|||   :::   :::
     686 AACAGTGGTCTCTCAGAGCCTGTCCAAAAACCCGGATCAGCCTCGTCCTC 735

     234 .....................SerTrpThrAspSerGluValAspSer 242
                                  :::   |||:::|||::::::   :::
     736 AGCCAGATCCTTATCAGATATTAGGACCCACCAGCAGTCGACTTGCTAAC 785

     243 SerCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGlnLeuLe 259
           |||||||||   |||:::||||||||||||||   ||||||||
     786 CCAGGCTCAGGTCAG...ATCCAGCTGTGGCAGTTCCTGCTGGAGCTCCT 832

     259 uLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGlnL 276
         |          :::   |||      |||   |||   :::      :
     833 CTCCGACAGCGCCAACGCCGGC...TGCATCACCTGG...GAAGGAACCA 876

     276 ysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGly 292
         ::|||   ||||||:::   |||      :::||||||||   ||||||
     877 ACGGAGAGTTCAAGATCGACGGACCCAGACAGAGGTGGCACGGCGCTGGGGC 926

     293 IleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIl 309
         ||||||:::::||:::|||||||||||||||||||||||||||:::::
     927 GAACGCAACAACAAGCCCAGCATCAACTACGACAAGCTCAGCCGCAGCCCT 976

     309 eArgGlnTyrTyrLysLysGlyIleIleArgLys 320
         :|||   ||||||   |||      :::   |||
     977 GCGCTACTACTACGACAAGACCACCATGACCAAA 1010

seq_name: gb_pr5:AF175387

seq_documentation_block:
LOCUS       AF175387     1547 bp    mRNA            PRI       14-FEB-2000
DEFINITION  Homo sapiens transcription factor ets (TEL2) mRNA, complete cds.
ACCESSION   AF175387
VERSION     AF175387.3  GI:6970607
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1547)
  AUTHORS   Potter,M.D., Buijs,A., Kreider,B. and Grosveld,G.
  TITLE     Cloning and characterization of a novel ETS gene, tel2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1547)
  AUTHORS   Potter,M.D., Buijs,A., Kreider,B. and Grosveld,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-1999) Genetics, St. Jude Children Research
            Hospital, 332 North Lauderdale, Memphis, TN 38105, USA
REFERENCE   3  (bases 1 to 1547)
  AUTHORS   Potter,M.D., Buijs,A., Kreider,B. and Grosveld,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Genetics, St. Jude Children Research
            Hospital, 332 North Lauderdale, Memphis, TN 38105, USA
  REMARK    Sequence update by submitter
COMMENT     On Feb 14, 2000 this sequence version replaced gi:6970212.
```

```
FEATURES             Location/Qualifiers
     source          1. .1547
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "6"
                     /map "6p21"
                     /tissue_type "hematopoietic"
     gene            1. .1547
                     /gene "TEL2"
     CDS             26. .1051
                     /gene "TEL2"
                     /note "similar to TEL1"
                     /codon_start 1
                     /product "transcription factor ets"
                     /protein_id "AAF25007.1"
                     /db_xref "GI:6970608"
                     /translation "MQEGELAISPISPVAAMPPLGTHVQAICEAQINLLGEGGICKLP
                     GRLRTQPALWSREDVLHWLRWAEQEYSLPCTAEHGFEMNGKALCTLTKDDFRHRAPSS
                     GDVLYELLQYIKTQKRALVCGPFFGGIFERLKTPTQHSPVPTEEKVTGPSQMDTRHGHLL
                     QPPDPGLTSNFGHLDDPGLARWTPGKEESLNLCHCAELGCRTGGVCSFPAMPQAPIDC
                     RTADCRLLWDYVYQLTLDTRYVFPYIEWEDEDAEIPPCVDENCLARLWCNHENEVEMTY
                     EEMSKALRHYYKLNITKKEPQQELLPRELKTPGSIVQDRGSHLSPLESQSQQSRTEEED
                     KRPETSP"
BASE COUNT       391 a    448 c    447 g    271 t
ORIGIN


alignment scores:
            Quality:  296.50              Length:     629
              Ratio:  1.742                 Gaps:       8
 Percent Similarity:  47.812   Percent Identity: 24.688

alignment block:
 US-09-126-945B-2 x AF175487  ..

 Align seq 1/1  to: AF175487  from: 1  to: 1547

       86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
          ||||||         |||         :::|||::       :::|||::
       17 CCACCTGTAATGCAGGAGGAGAATTGGCTATTTCTCCTATAAGCCCTGT 66

      102 aGlySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerL 119
          ::::::::::         |||   |||      ::::::
       67 GGCAGCCATGCCT.....CCCCTAGGCACCCACGTGCAAGCCACATGTG 110

      119 euGluGlnValGlnSerMetValValGlyGluValLeuLysAspIleGlu 135
             |||::::::   :::      |||   :::
      111 AAGCTCAAATTAACCTGCTGGGTGAAGGGGGGATC.............. 145

      136 ThrAlaCysLysLeuLeu...AsnIleThrAlaAspProMetAspTrpSe 151
             |||||||||       ::::::      :::|||       |||||
      146 .....TGCAAGCTGCCAGGAAGACTCCGCATCCAGCCCGCACTGTGGAC 189

      151 rProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgL 168
          |   :::::|||      ||||||   |||:::|||:::::::|||   |
      190 CAGGGAGGACGTGCTGCACTGGCTGCCTGGGCAGAGCAGGAGTACTCTC 239

      168 euProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCys 184
          |||||      ::::::          |||:::   |||:::   ||||||
      240 TGCCATGCACCGCGGAGCACGGGTTCGAGATGAACGGACGCGCCCTCTGC 289

      185 AlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspVa 201
          :::::::::::::|||||:::|||:::|||   :::||||||||
      290 ATCCTCACCAAGGACGACTTCCGGCACCGTGGCGCCCAGCTCAGGTGACGT 339

      201 lLeuHis..................................................... 203
          |||:::
      340 CCTGTATGAGCTGCTCCAGTACATCAAGACCCAGCGGCGAGCCCTGGTGT 389

      203 ..... 203

      390 GTGGGCCCTTTTTTGGAGGATCTTCAGGCTGAAGACGCCCACCCAGCAC 439

      203 ..... 203

      440 TCTCCAGTCCCCCCGGAAGAGGTGACTGGCCCCTCTCAGATGGACACCCG 489

      204 .....A 204
                :
      490 AAGGGGCCACCTGCTGCAGCCACCAGACCCAGGGCTTACCAGCAACTTCG 539

      204 IaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSer 220
          ::|||||||||          |||   |||      :::   :::
      540 GCCACCTGGATGACCCTGGCCTGGCAAGGTGGACCCCTGGCAAGGAGGAG 589

      221 ProGlyAlaIleHisTyrCysAlaSer..................... 229
                :::   :::|||||:::
      590 TCCCTCAACTTATGTCACTGTGCAGAGCTCGGCTGCAGGACCCAGGGGGT 639

      230 .ThrSerGluGluSerTrpThrAspSerGluValAspSerSerCysSerG 246
          |||      :::      ::::::   :::|||:::      :::
      640 CTGPTCCTTCCCCGCGATGCCGCAGGCCCCCATTGACGGCAGGATCGCTG 689

      247 lyGlnProIleHisLeuTrpGlnPheLeuLysGlnLeuLeuLysPro 262
                :::   |||||||||||||||   :::|||||||||
      690 ACTGCCGCCTG...CTGTGGGATTACGTGTATCAGCTGCTCCTTGATACC 736

      263 HisSerTyrGlyArgPheIleArgTrpLeuAsnLysGlnLysGlyIlePh 279
          :::   |||      :::|||:::|||   :::|||:::      |||||
      737 CGA...TATGAGCCCTACATCAAGTGGCAAGACAAGGACGCCAAGATCTT 783

      279 eLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysA 296
          |:::      |||          :::|||||||||||||||   :::|||
      784 CCGAGGTGTGGATCCAAATGGGCTCGCCAGACTCTGGGGAAATCACAAGA 833

      296 snArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyr 312
          |||||      |||:::|||:::|||:::|||||:::::|||:::|||
      834 ACGGGTTGGACATGACCTACGAGAAGATGTCTCGTGCCCTGCGCCACTAT 883

      313 TyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTy 329
          ||||||   :::|||||:::|||   :::   :::|||:::|||:::::
      884 TATAAGCTTAAATATCATTAAGAAG...GAACCGGGGCAGAAACTCCTGTT 930

      329 rGlnPheVal 332
          :::|||:::
      931 CAGATTTCTA 940

seq_name: qb pat1:AR104484

seq documentation block:
LOCUS        AR104484      5667 bp    DNA              PAT       14 FEB-2001
DEFINITION   Sequence 4 from patent US 6087485.
ACCESSION    AR104484
VERSION      AR104484.1  GI:12815072
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 5667)
  AUTHORS    Brooks Wilson,A.R., Buckler,A., Cardon,L., Carey,A.H., Galvin,M.,
             Miller,A. and North,M.
  TITLE      Asthma related genes
  JOURNAL    Patent: US 6087485-A 4 11 JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5667
                     /organism "unknown"
BASE COUNT     1662 a   1102 c   1226 g   1677 t
ORIGIN


alignment scores:
            Quality:  266.00              Length:     428
              Ratio:  1.298                 Gaps:       18
 Percent Similarity:  47.897   Percent Identity: 25.467

alignment block:
```

```
US-09-126-945B-2 x AR103484  ..

Align seq 1/1  to: AR103484  from: 1  to: 5667

      1 MetGlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLe 17
        :::|||:::  :::       :::  ||||||  :::|||:::||
     57 TTGGGAGGCATTGGCTCCACTGCAGCCTGGCTGTCTAGAGACCACATTCT 106

     17 u...LeuProProAspThrValSerArgThr.GlyLeuGluLysAlaAla 72
        |  ||||||          :::|||||| |||          :::
    107 CACCCTGCCTTTGTTACTGGGAAACCGAACGCGGCGCTGTGGCTTTCAGC 156

     33 AlaGlyAlaValGlyLeuGlu..........................Ar 40
           |||     ||||||:::                          ||
    157 TTGGGTAAGCCGGGTCTGCGGCGGGGATTGCCATCTGAAGACAGAGCCAG 206

     40 qArqAspTrpSerProSerProProAlaThrPro.............. 51
        ||||     :::|||  |||  |||  |||
    207 GAGGGCAGCCACACCTTGCCCAGCTGCACACCCAGTAACAAGTTTCCTCA 256

     52 ..............................GluGlnGlyLeuSerAla 67
                                     ::::::::::::::::
    257 GTGCGGGTATCTGCCACAGGCTGGGCTGGTCATCAAAGGGCCTCAGTCAT 306

     58 PheTyrLeuSerTyrPheAspMetLeu.............TyrProGluAs 70
        :::  ::::::  |||      |||           |||  |||
    307 ATTTAATAGAGCTCTTCAAGTATCTGGCTTTGTGATAATATCAGGAA.. 354

     73 pSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArqGluGluProP 87
                     :::  |||  |||:::|||:::
    355 ...................TCAGTTGGTTTCTCTGACAGACACTG..... 400

     87 roGluGluProGluGlnCysProValIleAspSerGlnAlaProAlaGly 103
                          |||:::|||           ::::|
    401 ...................CCCATTATCATGATTCTGGAAGGAGGTGGT 410

    104 SerLeuAspLeuValProGlyGlyLeuThrLeu.................. 114
        ::::::::|  ||||||:::      |||
    411 GTAATGAATCTCAACCCCGGCAACAACCTCCTTCACCAGCGCCAGCGTG 460

    115 ....GluGluHisSerLeuGluGlnValGlnSerMetValValGlyGlnV 130
            ::::::::::|||      :::|||:::|||         |||
    461 GACAGACAGCTACTCCACGTGCAATGTTTCCAGTGGCTTTTTGGAGCGC 510

    130 alLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAsp 146
               :::|||:::
    511 AGTGGCATGAAATTCAT................................ 527

    147 ProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGl 163
        |||:::  |||:::      :::|||  :::|||||
    528 CCTCAGTACTGGACCAACTACCAGCTCTGCGACTGCCTCCACACCTCCT 577

    163 uHisGlnTyrArqLeuProProMetGlyLysAlaPheGlnGlu.....L 178
              :::|||                 ||||||||||      :
    578 GGACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTTCGACA 627

    178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArqGlnArq 194
        ::  |||:::::|||||:::|||||  :::::|||  :::
    628 TCAACGCCAGCACCTCTCCACCATGACTTTCCACGACTTCAGCCGCCGC 677

    195 SerProLeuGlyGlyAspValLeuHisAlaHisLeu.............. 206
        :::      :::|||:::::|||:::::::|||
    678 GCAGCGACCGCGGGGCAGCTCCTCTACAGCAACTTGCAGCATCTGAACTG 727

    207 ...................AspIleTrpLysSer.........AlaAlaI 214
                           |||:::::::::|||      :::
    728 GAACGGCCAGTGCAGTACTGACCTGTTCCAGTCCACACACAATGTCATTC 777

    214 rpMetLysGluArgThrSerProGlyAlaIle................... 224
              |||:::|||:::|||:::  :::
    778 TCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAGACGAG 827

    225 HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGluValAs 241
        :::|||      :::|||:::      |||  :::|||      :::||
    828 AACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG...TTGGA 874

    241 pSerSer......CysSerGlyGlu........................... 247
        |||:::      |||  :::|||
    875 CAGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAGTCACC 924

    248 ..Proile.............................................. 249
          ||:::
    925 TTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCCCTGCC 974

    250 .....................................HisLeuTrpGlnPh 254
                                           ||||||||||:::||
    975 AAGTGCCACACCAAAAAGCACAACCCGAGAGGCACTCACTTATGGGAATT 1024

    254 eLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArqPheIleArqT 271
        |:::::::::::|||||:::|||  :::          :::||||:::|
   1025 CATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATAAAAT 1074

    271 rpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGluVal 287
        ||  :::::::::::::|||:::||::::::      |||      |||
   1075 GGGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGCAGTG 1124

    288 AlaArqLeuTrpGlyIleArqLysAsnArqProAlaMetAsnTyrAspLy 304
        |||:::||||||||  :::||||||:::  :::|||:::|||:::||
   1125 GCTCAGCTATGGGGTAAAAAGAACAACAACAGCAGCATGACCTATGAAAA 1174

    304 sLeuSerArqSerIleArqGlnTyrTyrLysLysGlyIleIleArqLysP 321
        ||||||||||:::::::|  ||||||||||:  ||||:::::::::
   1175 GCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATACTGGAGCGTG 1224

    321 roAspIleSerGlnArqLeuValTyrGluPhe 331
          |||  :::::|||||||||||||:::|||
   1225 TGGAT...GGACGAAGACTGGTATATAAATTT 1253

seq_name: qb_htq2:AC009471

seq_documentation_block:
LOCUS       AC009471   117175 bp    DNA             HTG       18-FEB-2000
DEFINITION  Drosophila melanogaster clone RPCI98-44K7, *** SEQUENCING IN
            PROGRESS ***, 69 unordered pieces.
ACCESSION   AC009471
VERSION     AC009471.5  GI:6996668
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 117175)
  AUTHORS   Muzny,D.M., Adams,C., Pailey,M., Barbaria,J., Blankenburg,K.,
            Bodoia,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunar,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Duqan-Rocha,S., Durbin,K.J., Fernandez,C., Ferrando,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,J.L.,
            Guevara,W., Harris,K., Hernandez,J., Hobinson,A., Hoques,M.,
            Holloway,C., Hosak,H., Jackson,L.F., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kmiejewski,N., Feng,Y., Fevar,C., Leal,R., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,B., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Snegant,K.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,L.K., Williamson,A.,
            Worley,K., Wren,J., Wronsford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 117175)
  AUTHORS   Worley,K.C.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (20 AUG 1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 18, 2000 this sequence version replaced gi:5881507.
                  Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
                  Project Information
            Center project name: DRAF
            Center clone name: RPC198 44K7
                  Summary Statistics
            Sequencing vector: M13; L08821
            Chemistry: Dye-primer Bodipy; 74% of reads
            Chemistry: Dye-terminator Big Dye; 25% of reads
            Assembly program: Phrap; version 0.980611
            Consensus quality: 58887 bases at least Q40
            Consensus quality: 78530 bases at least Q30
            Consensus quality: 86616 bases at least Q20
            Estimated insert size: 101921; sum-of-contigs estimation
            Quality coverage: 1x in Q20 bases, sum of contigs estimation
            - - - - - - - - -.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 69 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1      1361: contig of 1361 bp in length
          *     1362      1381: gap of unknown length
          *     1382      2346: contig of 965 bp in length
          *     2347      2366: gap of unknown length
          *     2367      3289: contig of 923 bp in length
          *     3290      3309: gap of unknown length
          *     3310      4086: contig of 777 bp in length
          *     4087      4106: gap of unknown length
          *     4107      5225: contig of 1119 bp in length
          *     5226      5245: gap of unknown length
          *     5246      6328: contig of 1083 bp in length
          *     6329      6348: gap of unknown length
          *     6349      7109: contig of 761 bp in length
          *     7110      7129: gap of unknown length
          *     7130      8387: contig of 1258 bp in length
          *     8388      8407: gap of unknown length
          *     8408      9836: contig of 1429 bp in length
          *     9837      9856: gap of unknown length
          *     9857     11360: contig of 1504 bp in length
          *    11361     11380: gap of unknown length
          *    11381     12234: contig of 854 bp in length
          *    12235     12254: gap of unknown length
          *    12255     13732: contig of 1478 bp in length
          *    13733     13752: gap of unknown length
          *    13753     14661: contig of 909 bp in length
          *    14662     14681: gap of unknown length
          *    14682     15665: contig of 984 bp in length
          *    15666     15685: gap of unknown length
          *    15686     16920: contig of 1235 bp in length
          *    16921     16940: gap of unknown length
          *    16941     17788: contig of 848 bp in length
          *    17789     17808: gap of unknown length
          *    17809     18722: contig of 914 bp in length
          *    18723     18742: gap of unknown length
          *    18743     19512: contig of 770 bp in length
          *    19513     19532: gap of unknown length
          *    19533     20676: contig of 1144 bp in length
          *    20677     20696: gap of unknown length
          *    20697     21501: contig of 805 bp in length
          *    21502     21521: gap of unknown length
          *    21522     22866: contig of 1345 bp in length
          *    22867     22886: gap of unknown length
          *    22887     24258: contig of 1372 bp in length
          *    24259     24278: gap of unknown length
          *    24279     26327: contig of 2049 bp in length
          *    26328     26347: gap of unknown length
          *    26348     27917: contig of 1570 bp in length
          *    27918     27937: gap of unknown length
          *    27938     28880: contig of 943 bp in length
          *    28881     28900: gap of unknown length
          *    28901     29807: contig of 907 bp in length
          *    29808     29827: gap of unknown length
          *    29828     31948: contig of 2121 bp in length
          *    31949     31968: gap of unknown length
          *    31969     33572: contig of 1604 bp in length
          *    33573     33592: gap of unknown length
          *    33593     34891: contig of 1299 bp in length
          *    34892     34911: gap of unknown length
          *    34912     36110: contig of 1199 bp in length
          *    36111     36130: gap of unknown length
          *    36131     37688: contig of 1558 bp in length
          *    37689     37708: gap of unknown length
          *    37709     38541: contig of 833 bp in length
          *    38542     38561: gap of unknown length
          *    38562     40169: contig of 1608 bp in length
          *    40170     40189: gap of unknown length
          *    40190     41344: contig of 1155 bp in length
          *    41345     41364: gap of unknown length
          *    41365     42596: contig of 1232 bp in length
          *    42597     42616: gap of unknown length
          *    42617     43451: contig of 835 bp in length
          *    43452     43471: gap of unknown length
          *    43472     44284: contig of 813 bp in length
          *    44285     44304: gap of unknown length
          *    44305     46132: contig of 1828 bp in length
          *    46133     46152: gap of unknown length
          *    46153     48272: contig of 2120 bp in length
          *    48273     48292: gap of unknown length
          *    48293     49747: contig of 1455 bp in length
          *    49748     49767: gap of unknown length
          *    49768     51323: contig of 1556 bp in length
          *    51324     51343: gap of unknown length
          *    51344     52829: contig of 1486 bp in length
          *    52830     52849: gap of unknown length
          *    52850     53927: contig of 1078 bp in length
          *    53928     53947: gap of unknown length
          *    53948     55808: contig of 1861 bp in length
          *    55809     55828: gap of unknown length
          *    55829     57498: contig of 1670 bp in length
          *    57499     57518: gap of unknown length
          *    57519     59461: contig of 1943 bp in length
          *    59462     59481: gap of unknown length
          *    59482     60737: contig of 1256 bp in length
          *    60738     60757: gap of unknown length
          *    60758     63029: contig of 2272 bp in length
          *    63030     63049: gap of unknown length
          *    63050     65083: contig of 2034 bp in length
          *    65084     65103: gap of unknown length
          *    65104     66948: contig of 1845 bp in length
          *    66949     66968: gap of unknown length
          *    66969     69391: contig of 2423 bp in length
          *    69392     69411: gap of unknown length
          *    69412     71348: contig of 1937 bp in length
          *    71349     71368: gap of unknown length
          *    71369     74204: contig of 2836 bp in length
          *    74205     74224: gap of unknown length
          *    74225     75348: contig of 1124 bp in length
          *    75349     75368: gap of unknown length
          *    75369     77168: contig of 1800 bp in length
          *    77169     77188: gap of unknown length
          *    77189     78329: contig of 1141 bp in length
          *    78330     78349: gap of unknown length
          *    78350     81411: contig of 3062 bp in length
          *    81412     81431: gap of unknown length
```

```
            *    81432    83284: contig of 1853 bp in length
            *    83285    83304: gap of unknown length
            *    83305    85285: contig of 1981 bp in length
            *    85286    85305: gap of unknown length
            *    85306    87622: contig of 2317 bp in length
            *    87623    87642: gap of unknown length
            *    87643    90091: contig of 2449 bp in length
            *    90092    90111: gap of unknown length
            *    90112    92885: contig of 2774 bp in length
            *    92886    92905: gap of unknown length
            *    92906    95275: contig of 2370 bp in length
            *    95276    95295: gap of unknown length
            *    95296    97811: contig of 2516 bp in length
            *    97812    97831: gap of unknown length
            *    97832   101213: contig of 3382 bp in length
            *   101214   101233: gap of unknown length
            *   101234   105790: contig of 4557 bp in length
            *   105791   105810: gap of unknown length
            *   105811   109096: contig of 3286 bp in length

alignment_scores:
            Quality:  266.00              Length:     942
              Ratio:  1.462                 Gaps:      14
 Percent Similarity:  53.216   Percent Identity: 29.522

alignment_block:
 US-09-126-945B-2 x AC009371   ..

 Align seq 1/1  to: AC009371  from: 1  to: 117175

      26 ThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgArgAs 42
         |||         :::|||:::|||   |||||||||          |||
   93532 ACAACTACTACGGAAGCAGTGGCGCCGGCAGTGGGGGCACCAGCCAGG.. 93579

      42 pTrpSerProSerProProAlaThrProGluGlnGlyLeuSerAlaPhe. 58
          ||||||   :::|||||||                  ::::|||:::
   93580 .TGGTCCGGGAAACCCCCA                  AACGCCTGGC 93607

      59 TyrLeu......SerTyrProAspMetLeuTyrProGluAspSerSerTr 78
         ||||||      ||||||||||:::
   93608 TACCTGACAACGTCCTACTTCAAC..................... 93631

      79 pAlaAlaLysAlaProGlyAlaSerSer.ArgGluGluProProGluGlu 89
                ||||||   |||:::::: ::::::::::|||   |||||
   93632 .........GCACCGACAGCGGCAACGGCAGCGGCGTCCCAGCGAGGAA 93671

      90 ProGluGlnCysProVal11e.AspSerGlnAlaProAlaGlySerLeuA 106
         ||:::::::      :::|||    :::   |||   |||:::::::
   93672 CCACAATCAATGGCTATCATTCCCTGCACCAGCAGCAGCAGCAGCAACAG 93721

     106 spLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnVal 122
           ::::::      :::   :::                       |||
   93722 CAGTCGCAGCAATCGCAGCAGCAACAGCAACTGGCCCACCAGCAACTGTC 93771

     123 GlnSerMetValValGlyGluValLeuLysAspIleGluThrAlaCysLy 139
               ::::::|||   :::      :::   :::::::::
   93772 GCACCAGCAGCAGCAGGCGTTGCATCAGCAACTGTCGCATCAACAGCAGC 93821

     139 sLeuLeuAsnIleThrAlaAsp.............ProMetAspTrpSerP 152
           :::   ||||||                 |||   :::   |||
   93822 AGCAGCAACAGCAACAGCAGCAACACCCGCACTCCCAACTGAACGGTCGG 93871

     152 roSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeu 168
           |||   :::   :::   |||       :::       :::   |||
   93872 CATCCGCACTCACATCCGCACTCTCATCCGCATTCGCATCCCCACGCGGG 93921

     169 ProProMetGlyLysAlaPheGlnGluAlaGlyLysGluGlyCysAl 185
         |||       :::   ||||||:::::   ::::::         ::::|
   93922 CCA......GCACACGCATTCCACAATCGCTGCAGCGGCGGCACGGCAGC 93966

     185 aMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspVal1 201
         |   |||:::::|||   :::|||:::          :::|||   :::|
   93966 AGCATCTAGTCGTCAGCAGCAGCAGCAGCGCCGTTGCAGCGGCACCATGC 94015

     202 euHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArg 218
         ||   |||                 ||||||||   ::::::::::|||
   94016 TGAGCGCC.................AGTGCAGCGGCACAGCAACGGCGG 94047

     219 ThrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGlu..... 233
               |||   ||:::          :::   :::   :::
   94048 CACTGCGGCGGGCGGATCCAGAGTGTTATCCAACCGGCGACGTCCAGCGT 94097

     234 .........................SerTrpThrAspSerGlu.ValAsp 241
                                 ||||||:::|||||:::: :::::
   94098 CACTACGATCTCTCCTACATGCTGGAGCTGGGCGGATTCCAGCAGCGGAA 94147

     242 ..............................................SerSe 243
                                                      :::
   94148 GGCGAAGAAGCCGCGCAAACCGAAGCTGGAGATGGGCGTAAAGCGGCGCA 94197

     243 rCysSerGlyGlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuL 260
           :::|||:::      :::|||||||:::||||||   :::|||||||
   94198 CAGTGAGGCATCCACCACCTATCTGTGGGAGTTCCTCCTCAAACTGCTCC 94247

     260 eLysProHisSerTyrGlyArgPheIleArgProLeuAsnLysGluLys 276
               :::          |||:|||||:::      ||:::|||||
   94218 AGCATCCGCAATACTGTCCCCCTTCATCAAGTACACAAACCCGCACAAC 94297

     277 GlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIl 293
         |||:::|||||||:::   ||||||      |||:::|||||||||||:::
   94298 GGCGTCTTCAAGCTGGTCGACTCGAAGGCTGTGTCCCGCCTGTGGGGCAT 94347

     293 eArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleA 310
         ::::|||||:::|||   |||||||||||:::   ::::::|||:::::|
   94348 GCACAACAACAAGCCGCACATGAACTACGACAAGCTGAGCAGGGCGCTGA 94397

     310 rqGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArg 326
         ||   ||||||:::::|||||:::   |||   |||   :::|||||||
   94398 GGTACTACTACCAGCGCGGCATACTGGCCAAGGTGGAT...GGCCAGCGG 94444

     327 LeuValTyrGlnPheVal 332
         |||||||||||||||||
   94445 CTGGTCTACCAGTTCGTG 94462

seq_name: qb_pri10:HSU73844

seq_documentation_block:
LOCUS       HSU73844     1846 bp    mRNA            PRI       06-AUG-1997
DEFINITION  Homo sapiens epithelial-specific transcription factor ESE-1a
            (ESE 1) mRNA, complete cds.
ACCESSION   U73844
VERSION     U73844.1  GI:1841524
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1846)
  AUTHORS   Oettgen,P., Alani,R.M., Barcinski,M.A., Brown,L., Akbarali,Y.,
            Boltax,J., Kunsch,C., Munger,K. and Libermann,T.A.
  TITLE     Isolation and characterization of a novel epithelium-specific
            transcription factor, ESE-1, a member of the ets family
  JOURNAL   Mol. Cell. Biol. 17 (8), 4419-4433 (1997)
  MEDLINE   97378019
REFERENCE   2  (bases 1 to 1846)
  AUTHORS   Oettgen,P., Akbarali,Y., Barcinski,J., Boltax,J., Kunsch,C. and
            Libermann,T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-1996) Medicine, Beth Israel Hospital, 330
            Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1846
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="adult liver, pancreatic carcinoma"
                     /chromosome="1"
```

```
                     /map "1q42.1"
     gene            1..1846
                     /gene "ESE-1"
     CDS             120..1166
                     /gene "ESE-1"
                     /note "alternative splice form of ESE-1,
                     epithelial specific; member of the Ets transcription
                     factor family"
                     /codon_start=1
                     /product "ESE-1a"
                     /protein_id "AAB65824.1"
                     /db_xref "GI:1841525"
                     /translation "MAATCEISNIFSNYFSAMYSSEDSTLASVPPAATFGADDLVELL
                     SNPQMSLEGTEKASWTGEQPQFWSKIQVLDWISYQVEKNKYDASAIDFSRCDMDGATL
                     CNCALEELRLVFGPLGDQLHAQLRDLTSSSSDELSWIIELLEKDGMAFQEALDPGPFD
                     QGSPFAQELLDDVSTAGTGASRSSHSSDSGGSDVDLDPTDGKLFPSDGFRDCKKGDPK
                     HGKPKPCPPPKLSKEYWFCLPSKKSKHAPECTHLWFIFGCLIHPELNPCLMEWENKH
                     EGVFKFLRSEAVAQLWGQKKKNSNMTYEKLSRAMRYYYKRETLERVDGRRLVYKFGKN
                     SSGWKEEEVLQSRN"
BASE COUNT      403 a    534 c    529 g    380 t
ORIGIN


alignment scores:
            Quality:  265.50              Length:    412
              Ratio:  1.420                 Gaps:     18
 Percent Similarity:  45.388   Percent Identity: 24.515

alignment block:
 US-09-126-945B-2 x HSU74844   ..

 Align seq 1/1  to: HSU74844  from: 1  to: 1846

      25 ArqThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArqAr 41
         |||:::|||      :::|||||||||      ::::::          ||
       3 AGAGCCGGGTAGGGGAGCGCAGCGGCCAGATACCTCAGCGCTACCTGGCG 52

      41 qAspTrpSerProSerProPro.......................... 48
         |:::|||         ||||||
      53 GAACTGGATTTCTCTCCCGCCTGCCGGCCTGCCTGCCACAGCCGGACTCC 102

      49 .....................AlaThrProGluGlnGlyLeuSerAla 57
                              ||||||   |||      :::|||
     103 GCCACTCCGGTAGCCTCATGGCTGCAACCTGTGAG......ATTAGCAAC 146

      58 PheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAl 74
         ::::::   :::||||||:::   :::|||   ::::::::::|||
     147 ATTTTTAGCAACTACTTCAGTGCGATGTACAGCTCGGAGGACTCCTCCCT 196

      74 qAlaLysAlaProGlyAlaSerSerArqGluGluProProGluGluProG 91
         |||:::::::|||   |||
     197 GGCCTCTGTTCCCCCTGCT............................... 215

      91 luGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeu 107
                                 |||      ||||::   ||||||
     216 ........................GCCACCTTTGGGGCCGATGACTTG 239

     108 ValProGlyGlyLeuThrLeuGlu......GluHisSerLeuGluGlnVa 122
         |||          ||||||||||:::      :::   ||||||||||   ::
     240 GTA.........CTGACCCTGAGCAACCCCCAGATGTCATTGGAGGGTAC 280

     122 lGln...SerMetValValGlyGluValLeuLysAspIleGluThrAlaC 138
         ::::   :::      :::||||||
     281 AGAGAAGCCCAGCTGGTTGGGGGAA....................... 305

     139 ysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsn 154
                                 :::||||::   ||||||   ::::::
     306 ......................CAGCCCCAGTTCTGGTCGAAGACGCAG 332

     155 ValGlnLysTrpLeuLeuTrp......ThrGluHisGlnTyrArqLeuPr 169
         |||      ||::::   :::          ::::::::|||
     333 GTTCTGGACTGGATCAGCTACCAAGTGGAGAAGAACAAGTACGACGCAAG 382

     169 oProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaM 186
         :::      :::      ::::::   |||      ||||||
     383 CGCCATTGACTTCTCACGATGTGACATGGATGGCGCACCTTCTGCAATT 432

     186 etSerGluGluGlnPheArq...GlnArqSerProLeuGlyGlyAspVal 201
         :::   |||:::::::|||          :::||||||   ||||||
     433 GTGCCCTTGAGGAGCTGCGTCTGGTCTTTGGGCCCTCTG...GGGGACCAA 479

     202 LeuHisAlaHistoG   AspIleTrpLysSerAla            Al 214
         ||||||||||::|||   |||:::   :::|||:::           ::
     480 CTCCATGCCCAGCTGCGAGACCTCACTTCCAGCTCTTCTGATGAGCTCAG 529

     214 aTrpMet.................................LysGluArqI 219
         :|||:::                               :::|||
     530 TTGGATCATTGAGCTGCTGGAGAAGGATGGCATGGCCTTCCAGGAGGCCC 579

     219 hrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluSerTrp 235
         :::||||||   :::          :::      :::::::|||
     580 TAGACCCAGGGCCCTTGACCAGGGCAGCCCCTTTGCCCAGGAGCTGCTG 629

     236 ThrAsp..................................................... 237
         |||
     640 GACGACGTCTCCACCGCAGGGACTGGTGCTTGTCGGAGCTCCCACTCCTC 689

     238 ............SerGluValAsp.......................... 241
                     |||:::||||||
     690 AGACTCCGGTGGAAGTGACGTGGACCTGGATCCCACTGATGGCAAGCTCT 739

     242 ................SerSerCysSer...................... 245
                         :::|||:::
     740 TCCCCAGCGATGCTTTCGTGACTGCAAGAAGGGGGATCCCAAGCACGGG 789

     246 ............GlyGlnPro............................. 248
                     |||:::|||
     790 AAGCGGAAACGAGGCCGGCCCCGAAAGCTGAGCAAAGAGTACTGGGACTG 839

     249 ....................................IleHisLeuTrpG 253
                                            |||||||||:
     840 TCTCGACGGCAAGAAGAGCAAGCACGCGCCCAGAGGCACCCACCTGTGGG 889

     253 lnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArqPheIle 269
         ::|||:::::::::::::|||:::   |||:::              ::::::
     890 AGTTCATCCGGGACATCCTCATCCACCCGCAGCTCAACGAGGGCCTCATG 939

     270 ArqTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 286
         :::|||   |||:::::::::|||:::|||||||:::      |||
     940 AAGTGGGAGAATCGGCATGAAGGCGTCTTCAAGTTCCTGCGCTCCGAGGC 989

     286 nValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMetAsnTyrA 303
         ||||||:::|||||||||   :::|||:::::      |||:::|||:
     990 TGTGGCCCAACTATGGGGCCAAAAGAAAAAGAACAGCAACATGACCTACG 1039

     304 spLysLeuSerArqSerIleArqGlnTyrTyrLysLysGlyIleIleArq 319
         ::|||||||||||||:::::|||   ||||||||||:::   |||::::::
    1040 AGAAGCTGAGCCGGGCCATGAGGTACTACTACAAACGGGAGATCCTGGAA 1089

     320 LysProAspIleSerGlnArqLeuValTyrGlnPhe 331
         :::   |||   :::::|||||||||||||:::|||
    1090 CGGGTGGAT...GGCCGGCGACTCGTCTACAAGTTT 1112

seq_name: qb_pr5:AF124438

seq_documentation block:
LOCUS       AF124438     2517 bp    mRNA            PRI       27-JAN 2000
DEFINITION  Homo sapiens transcription factor ESE-3A (ESE-3) mRNA, complete
            cds.
ACCESSION   AF124438
VERSION     AF124438.1  GI:4838431
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 2517)
  AUTHORS   Kas,E., Finger,E., Grail,F., Gu,X., Akbarali,Y., Beltaa,J.,
            Weiss,A., Oettgen,P., Kapeller,R. and Libermann,T.A.
  TITLE     ESE-3, a novel member of an epithelium-specific ets transcription
            factor subfamily, demonstrates different target gene specificity
            from ESE-1
  JOURNAL   J. Biol. Chem. 275 (4), 2986-2998 (2000)
  MEDLINE   20112875
REFERENCE   2  (bases 1 to 2517)
  AUTHORS   Kas,E., Finger,E., Grail,F., Gu,X., Akbarali,Y., Beltaa,J.,
            Oettgen,P., Kapeller,R. and Libermann,T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28 JAN 1999) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..2517
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p14"
                     /tissue_type="prostate gland"
     gene            1..2517
                     /gene="ESE-3"
     CDS             98..931
                     /gene="ESE-3"
                     /note="Ets transcription factor"
                     /codon_start=1
                     /product="transcription factor ESE-3A"
                     /protein_id="AAD30990.1"
                     /db_xref="GI:4884842"
                     /translation="MILEGGVMNLNFGNNLLHQFFAWTLATSIFNVSSGFFGSQWHE
                     IHPQYWTKYQVWEWDQHLLDTNQLDANCIPFQEFDINGFDIQSMSIQEFTRAACTAGQ
                     LLYSNFQHLKWNGQCSSDLPQSFHNVIVKTFQFFPSTMNIWKHQNYLYDINYGSIVAE
                     SPDMKKFQDPPAKFTEFEHNFGTHGWFIEFGLLSGCKRHQTLFEEFSFQVRGLVYEPC
                     SEAVAQLWGKKKNNSSMTYEKLSPAMRYYYFRPETLEPVIGRRLVYEPVCENARGWPENE
                     N"
BASE COUNT       778 a    513 c    559 g    667 t
ORIGIN


alignment_scores:
            Quality:  265.50              Length:     329
              Ratio:  1.544                 Gaps:       9
 Percent Similarity:  52.280   Percent Identity: 24.924

alignment_block:
 US-09-126-945B-2 x AF124438  ..

 Align seq 1/1  to: AF124438  from: 1  to: 2517

       38 LeuGluArgArgAspTrpSerProSerProProAlaThrProGluGlnGl 54
          |||  |||  :::  :::  ||||||||||::|||          ||
       14 CTGGTGAGATTGCAGCCTGCTGCCTCCCCTCCATCAACCACAGCTATTGG 63

       54 yLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspS 71
          |:::  :::     |||          :::::::::     |||  :
       64 ATTTCCCACCCAGAATCTTTAGGTAAATGAGATCAICATGATTCTGGAAGGAG 113

       71 erSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluProPro 87
          :::::                |||||   :::      ::::::||||||
      114 GTGGTGTAATGAATCTCAACCCCGGCAACAACCTCCTTCACCAGCCGCCA 163

       88 GluGluProGluGlnCysProValIleAspSerGlnAlaProAlaGlySe 104

      163 ..................................................  163

      104 rLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGlnG 121
                                    ::::::::::|||      :
      164 ....................GCCTGGACAGACAGCTACTCCACATGCA 191

      121 lnValGlnSerMetValValGlyGluValLeuLysAspIleGluThrAla 137
          ::|||:::|||         |||:::         :::|||:::
      192 ATGTTCCAGCGCTTTTTTTGGAAGCCAGTGGCATGAAAATCCAC..... 235

      138 CysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAs 154
                                        ||||:::   ||||:::     ::
      236 .........................CCTCAGTACTGGACCAAGTACCA 258

      154 nValGlnLysTrpLeuLeuTrpIhrGluHisGlnTyrArgLeuProProM 171
          :|||   :::|||||||                    ::::||
      259 GGTGTGGGAGTGGCTCCAGCACCTCCTGGACACCAACCAGCTGGATGCCA 308

      172 ::GlyLysAlaPheProGlu      ::..AlaGlyLysGlu::CysAla 185
                  |||||||||         :::   ||:::::::||||:::
      309 ATTGTATCCCTTTCCAAGAGTTCGACATCAACGGCGAGCACCTTTGCAGC 358

      186 MetSerGluGlnGlnPheArgGlnArgSerProLeuGlyGlyAspValIe 202
          ||||||   :::::::||   :::   :::        :::|||:::::||
      359 ATGAGTTTGCAGGAGTTCACCCGGCGGCAGGGACGGCGGGGCAGCTCCT 408

      202 uHisAlaHisLeu.................................AspI 208
          |:::::::::::||                                ||||
      409 CTACAGCAACTTGCAGCATCTGAAGTGGAACGGCCAGTGCAGTAGTGACC 458

      208 leTrpLysSer.........AlaAlaTrpMetLysGluArgThrSerPro 221
          ::::::::::|         :::           |||:::|||:::||
      459 TGTTCCAGTCCACACACAATGTCATTGTCAAGACTGAACAAACTGAGCCT 508

      222 GlyAlaIle.....................................HisTy 226
          :::   :::                                    :::||
      509 TCCATCATGAACACCTGGAAAGACAGAGAACTATTTATATGACACCAACTA 558

      226 rCysAlaSerThrSerGluGluSerTrpThrAspSerGluValAsp.... 241
          |  ::::::::::::::::::                ::::||   |||
      559 TGGTAGCACAGTAGCAGACTCACCTGATATGAAAAAACAGCAAGACCCC 608

      [illegible]
          :::::::||          :::   :::      ||||||||
      609 CTGCCAAGTGCCACACCAAAAAGCACAACCCGACAGGCACTCACTTATGG 658

      253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl 269
          :::|||:::::::::::::|||||::::||   :::          ::||
      659 GAATTCATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAAT 708

      269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaC 286
          |::|||   :::::::::::::|||:::::::::      |||
      709 AAAATGGGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGG 758

      286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
          ||||||:::|||||||||   :::|||||:::   :::|||:::|||
      759 CAGTGGCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTAT 808

      303 AspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 319
          :::|||||||||||||:::::::|||   ||||||||||:::   |||::::::
      809 GAAAAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATTCTGGA 858

      319 gLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          ::::   |||   :::::|||||||||||||:::|||
      859 GCGTGTGGAT...GGACGAAGACTGGTATATAAATTT 892

seq_name: qb_ov:DR084615

seq_documentation_block:
LOCUS       DR084615     4200 bp    mRNA            VRT       14-JAN-1999
DEFINITION  Danio rerio E74 like factor 1 (ZEF1) mRNA, complete cds.
ACCESSION   U84615
VERSION     U84615.1  GI:4099173
KEYWORDS    .
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 4200)
  AUTHORS   Uchida,H., Zon,L., Miyazaki,Y. and Nimer,S.
  TITLE     Cloning of several members of the ETS family of proteins from
```

```
            zebrafish
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4200)
  AUTHORS   Uchida,H., Zon,L., Miyazaki,Y. and Nimer,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-1997) Hematologic oncology, Memorial
            Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY
            10021, USA
FEATURES             Location/Qualifiers
     source          1. .4200
                     /organism "Danio rerio"
                     /db_xref "taxon:7955"
                     /tissue_type "kidney"
     gene            1. .4200
                     /gene "ZEF1"
     CDS             458. .1665
                     /gene "ZEF1"
                     /codon_start 1
                     /product "E74-like factor-1"
                     /protein_id "AAD04753.1"
                     /db_xref "GI:4099174"
                     /translation="MTTVVQANFLVFFFASNGMDEINQLDDPSVFPAVTVEQVPAADL
                     MQVYSGLEADEVTNGIMVDAIQDVVEENMMGDVGLSVETPVSGGEDNMETIEAAEALL
                     NMESPNNILDEKRMIHTYGNILETDLSYISLRPEHLSNGMDVAHDDETSSLDEVPQKN
                     FSKPARKIKVRKPPPARPCSPTTNPSLPLKKKSKEGKCNFIYLWEFLLALLQDKNTCP
                     KYIKWTQREKGIFKLVDSKAVSKLWGKHKNKPDMNYEKMGKALRYYYQRGILAKVEGQ
                     RLVYQFKEMPTDLVIIDEDDGQCDDCGACYCCHPAAPRCPVCPGSSRSQCPNTQVKTV
                     KREMSPGLNSKERKTSKTASADGSCPPAKSRSCCPSADILRKDYKYARLCSHGPHVES
                     SRRRSNDPADPAAVLCPDQRGFLDSRLPTTGHTPHSAFNQPWR"
BASE COUNT     1173 a    974 c    979 g    1074 t
ORIGIN


alignment_scores:
            Quality:  265.50              Length:    279
              Ratio:  1.702                 Gaps:     12
 Percent Similarity:  55.914   Percent Identity: 30.108

alignment block:
 US-09-126-945B-2 x DRU84615   ..

 Align seq 1/1  to: DRU84615  from: 1  to: 4200

      88 GluGluProGluGlnCysPro...ValIleAspSerGlnAlaProAlaGl 103
         ::::::|||:::      |||   ||||||   :::|||:::||||||::
     433 GACGACCCTTCGGTGTTTCCAGCGGTGATTGTGGAGCAGGTGCCCGCCGC 482

     103 ySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSer.... 118
         :::|||      |||   :::||||||      :::|||   :::
     483 CCACCTCATGCAGGTGTACTCTGGGCTGGAGCCAGACGAGGTGACTAATG 532

     119 ........LeuGluGlnValGlnSerMetVal.............ValGly 128
                 ::::::   :::|||:::::|||             :::|||
     533 GCATCATGGTGGACGCCATTCAGGACGTGGTGGAAGAGAACATGATGGGA 582

     129 GluVal.....................................Le 131
         :::|||                                    ::
     583 GACGTGGGGCTGTCTGTAGAGACCCCAGTCTCAGGGGAGAAGACAACAT 632

     131 qLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProM 148
         ::::   ||||||::|||      ||||||||||:::
     633 GGAGACCATCGAGGCGGCTGAAGCCCTGCTGAACATGGAA.......... 672

     148 etAspTrpSerProSerAsnVal......GlnLysTrpLeuLeuTrpThr 162
                 ||||||::|||:::      ::|||   ::::::   |||
     673 .......TCTCCCAATAACATCCTGGATGAGAAGCGCATGATTCACACA 714

     163 GluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGluLeuAl 179
                                          :::   ::::::
     715 ..................................TATGGGAATATCCT 728

     179 aGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgSerP 196
               :::|||      :::|||   :::        ::::::
     729 GGAAACAGACCTGAGCTACATCTCCCTCCGTCCA......GAGCATCTCT 772

     196 roLeuGlyGlyAspValLeuHis.............AlaHisLeuAsp... 207
            |||   ||||||   |||              :::   ||||||
     773 CAAATGGCATGGATCTTGCCCACGATGATGAGACGTCCTCTTTGGATGAA 822

     208 ................IleTrpLysSerAlaAlaTrpMetLysGlnArgPh 219
                         :::   |||   |||          |||   |||
     823 GTCCCTCAGAAAAACTTTTCCAAGCCTGCCAGGAAAACCAAAGTGCGC.. 870

     219 rSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpT 236
          :::|||          :::   |||:::      ::::::   |||
     871 .AAACCAAGACCAGCGCGGCCATGCTCTCCAATAACCAATCCATCTCTCC 919

     236 hrAspSerGluValAspSerSerCysSerGlyGluProIleHisLeuTrp 252
               ::::::  ::::::::::  :::|||:::  ||||:::||||||
     920 CGCTGAAGAAGAAAAGCAAAGAGGGCAAAGGTAACACCATCTACCTGTGG 969

     253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl 269
         :::|||||      ||||||          ::::::      :::::::||
     970 GAGTTCCTGCTGCCCTTGCTGCAAGACAAAAACACCTGTCCCAAATACAT 1019

     269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGlyAspSerAlaG 286
         |:::|||   ::::::|||||||||||||||||||:::   ||||||
    1020 CAAGTGGACGCAGAGAGAAGGGCATTTTTAAGCTGGTGGACTCCAAGG 1069

     286 InValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
           |||::::::|||||||||   :::|||||||:::|||   |||||||||
    1070 CTGTCTCCAAACTGTGGGCAAAGCACAAAAACAAGCCTGATATGAACTAT 1119

     303 AspLysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleIleAr 319
         :::   ::::::|||:::::|||   ||||||:::::|||||||:::
    1120 GAAACCATGGGCCGAGCACTCAGGTACTATTACCAGAGGGGTATATTGGC 1169

     319 qLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          |||   :::   :::|||||||||||||||||||||
    1170 CAAACTGCAA...GCACACGCGCTTGTTTACCAGTTT 1203


seq_name: qb_ov:GGETSC

seq_documentation block:
LOCUS       GGETSC       1856 bp    mRNA            VRT       12-SEP-1993
DEFINITION  Chicken mRNA for c-ets proto-oncogene.
ACCESSION   X13027
VERSION     X13027.1  GI:63382
KEYWORDS    ets cellular oncogene; ets oncogene; proto-oncogene.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1856)
  AUTHORS   Chen,J.H.
  TITLE     Complementary DNA clones of chicken proto-oncogene c-ets: sequence
            divergence from the viral oncogene v-ets
  JOURNAL   Oncogene Res. 2 (4), 371-384 (1988)
  MEDLINE   88289029
COMMENT     See <X00144> for v-ets sequence of Avian erythroblastosis virus
            tragment E26.
FEATURES             location/Qualifiers
     source          1. .1856
                     /organism "Gallus gallus"
                     /db_xref="taxon:9031"
                     /tissue_type="spleen"
                     /clone_lib "lambda gt10"
     CDS             176. .1501
                     /note="c-ets gene product 50kD protein (AA 1-441)"
                     /codon_start 1
                     /protein_id="CAA31442.1"
                     /db_xref="GI:63383"
                     /db_xref "SWISS-PROT:P13474"
                     /translation="MKAAVDLKPTLTIKIPKVDIDLEFPSPLMECADVPLTTPSSKEM
                     MSQALKATFSGFAKEQQPLGIPKDPQQWTFTHVPLQWVMWAVNPFSLKGVDPQKFCMNG
```

```
                    AALCALGKFCPLFLRPDFVGDTLWFRLFTLQKFFAKPYPANGVNAAYPESRYFSDYFI
                    SYGIFHAQCVPPSFSEPSEPSFLFSVQTLFGGLSSFGTLSRLYNDSIPSVFLRDFVQSLS
                    LQTDYFTIKQEVVIPGNMCMGRASSGKLGGQQDSFFSSLESYDSCDRLIQSWSSQSSPQS
                    LQRVPSYSSFTLSFEAYFAALFNHPFFSTPKIYVFGPALMNGLFFFVCLGAACAAYLLRP
                    IQIWQFLIFLLTIGKSPQSPFISWTGIGWPFKLSDPDFVAPPWCFFFNSLFFMNYFFLSRG
                    LRYYYDRNTIHSIAGRSVYREVCDLQSLLGYIPSFLMAMLIVRFLAGR"
BASE COUNT      480 a    489 c    507 g    380 t
ORIGIN


alignment_scores:
            Quality:  264.50              Length:    446
              Ratio:  1.503                 Gaps:     17
 Percent Similarity:  39.462   Percent Identity: 25.318

alignment_block:
 US-09-126-945B-2 x GGEISC    ..

 Align seq 1/1  to: GGEISC  from: 1  to: 1856

      74 AlaAlaLysAlaProGlyAlaSerSerArgGluGluProProGluGluPr 90
         :::|||:::::|||||          ::::::   ||||||||||::
     107 ICGGCCCGCTGCCCCGGC.....CAGCAGCCGCCCCCCAGTTCCCA 147

      90 oGluGlnCysProValIleAspSerGlnAlaProAlaGly..... 103
         ::::::   |||         |||         ||::::
     148 GTCCCGGTCCCCTCGCTCCGGCTCAACCATGAAGGCGGCGGTGGACCTGA 197

     104 ..................................SerLeuAspLeuVal 108
                                          ::::::||||||
     198 AGCCCACCCTGACCATCATCAAGACGGAGAAGGTGGACATCGATCTCTTC 247

     109 Pro.................GlyGlyLeuThrLeuGluGluHisSerLe 119
         |||                ::: :::   |||         |||
     248 CCGTCCCCCGATATGGAATGTGCAGATGTGCCTTTGTTAACCCCCAGCAG 297

     119 uGluGlnValGlnSerMetValValGlyGluValLeuLysAspIleGluI 136
         ::::::::::|||:::::::    ::::::::   :::
     298 CAAGGAAATGATGTCTCAGGCACTGAAAGCCACCTTCAGTGGCTTCGCAA 347

     136 hrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerPro 152
             :::   |||::::|||      |||||||::::::|||:::
     348 AGGAGCAGCAGCGGCTGGGAATCCCCAAAGATCCCCAGCAGTGGACAGAG 397

     153 SerAsnValGlnLysTrpLeuLeuTrpIhrGluHisGlnTyrArgLeuPr 169
         ::::::|||:::   |||:::::::|||:   :::::::::   |||
     398 ACGCACGTGCGGGACTGGGTGATGTGGCCAGTGAATGAGTTCAGCCTG.. 445

     169 oProMetGlyLysAlaPheGlnGlu......LeuAlaGlyLysGluLeuC 184
             |||       ||||||||::      :::   |||      ||||
     446 ....AAGGGAGTGGATTTCCAGAAGTTCTGCATGAACGGAGCTGCCCTGT 491

     184 ysAlaMetSerGluGluGlnPheArgGlnArgSerPro...LeuGlyGly 199
         ||||||||||||||||   |||   :::      |||   :::   |||
     492 GCGGCCTGGGCAAGGAGTGCTTCCTGGAGCTACGGCCTGACTTTGTGGGA 541

     200 AspValLeuHisAlaHisLeuAspIleTrpLys..... 210
         |||::::||      ||||||::::|||   :::
     542 GATATCCTTTGGGAACACCTGGAGATCTTGCAGAAAGAAGAGGCAAAACC 591

     211 ..................SerAlaAlaTrpMetLysGluArg...IhrS 229
                         :::||||||:::   :::::::|||   |:||
     592 ATACCCAGCAAATGGAGTGAATGCAGCGTATCCAGAATCCCGCTATACTT 641

     220 erProGlyAlaIleHisTyr.................CysAlaSerIhr 236
         ||       |||   |||                 ||::::
     642 CAGACTACTTCATTAGTTATGGCATCGAGCACGCACAGTGCGTGCCTCCC 691

     231 SerGlu.....................GluSerTrp........... 245
         ||||||                     |||||||:::
     692 TCCGAGTTCTCTGAGCCCAGCTTCATCACAGAGTCCTACCAGACCCTCCA 741

     245 ..................................................... 245
```

```
     742 TCCCATCAGCTCGGAAGACCTTCTGTCCTCCAAGTACGAGAACGACTATC 791

     246 .............................IhrAspSer............ 258
                                      ||||||||
     792 CCTCAGTCATCCTTCGTGACCCCGTCCAGACGGACTCCCTGCAGACAGAC 841

     259 ..................GluVal.........AspSerSerCysSerGl 246
                          ||||||         ||::::   |||   ||
     842 TACTTCACAATCAAGCAAGAAGTGGTAACGCCAGACAACATGTGCATGGG 891

     246 yGln.................................................. 247
         |:::
     892 ACGTGCCAGTCGAGGTAAACTGGGTGGCAGGACTCCTTTGAGAGCATAG 941

     247 ..................................................... 247

     942 AGAGCTACGACAGCTGTGACCGCCTGACACAGTCCTGGAGCAGCCAGTCC 991

     247 ..................................................... 247

     992 TCCTTCCAGAGCCTGCAGCGCGTCCCCTCCTACGATAGCTTTGACTCAGA 1041

     247 ..................................................... 247

    1042 GGACTACCCCGCCGCCCTGCCCAACCACAAGCCCAAGGGCACCTTCAAGG 1091

     247 ..................................................... 247

    1092 ACTATGTTCGAGATCGGCCTGACATCAACAAGCACAAGCCTGTCATTCCT 1141

     248 ..............................ProIleHisLeuTrpGl 258
                                      ||||||::|||||||
    1142 GCCGCTGCCCTCGCCGGCTACACAGGCAGTGGACCCATCCAACTGTGGCA 1191

     258 nPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleA 270
         |||||||   |||   ||||||          |||          ||||||
    1192 ATTCCTGCTGGAG...CTGCTCACTGACAAGTCCTGTCAGTCCTTCATAA 1238

     270 rgTrpLeuAsnLysGlnGlyIleLysIleGluAspSerAlaGln 286
         |||   :::   :::                ||||::::::|||     :::
    1239 GCTGGACGGGT...GATGGCTGGCAGTTCAAGCTTTCCGATCCAGATGAG 1285

     287 ValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAs 303
         |||||||||   ||||||   ||||||||||::|||   |||||||||::
    1286 GTGGCCAGGCGCTGGGAGAAGAGGAAAAACAAGCCCAAGATGAACTATGA 1335

     303 pLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAraL 320
         :|||||||||||||:::::|||   ||||||   |||:::|||||||::|
    1336 GAAGCTGAGCCGTGGTCTGCGTTACTATTACGACAAGAACATCATCCACA 1385

     320 ysProAspIleSerGlnArgLeuValTyrGluPheVal 332
         ||        ::::::|||   ||||||::||||||
    1386 AG...ACGGCCGGCAAGCGCTACGTCTACCGCTTCGTC 1420

seq_name: qb_rol:BC005486

seq_documentation_block:
LOCUS       BC005486     3496 bp    mRNA            ROD       03-APR-2001
DEFINITION  Mus musculus, E26 avian leukemia oncogene 2, 3' domain, clone
            MGC:7318, mRNA, complete cds.
ACCESSION   BC005486
VERSION     BC005486.1  GI:13529535
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3496)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
```

```
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
   REMARK   NIH MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Lothar Bennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www.shgc.stanford.edu
            Contact:  (Dickson, Mark) med@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 7 Row: j Column: 9.
FEATURES             Location/Qualifiers
     source          1..3496
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="MGC:7418 IMAGE:3486021"
                     /tissue_type="Mammary tumor, WAP TGF alpha model, 7 months
                     old, gross tissue."
                     /clone_lib="NCI_CGAP_Mam5"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             160..1566
                     /codon_start=1
                     /product="E26 avian leukemia oncogene 2, 3' domain"
                     /protein_id="AAH05486.1"
                     /db_xref="GI:13529536"
                     /translation="MNDFCIKNMDQVAPVANSFRGTLKRQPAEDTFDGSLFAVLPSLS
                     EDQTLQEVPTGLDSVSHDSASCELPLTPCSKAVMSQALKATESGFQKEQRRLGIPKN
                     PWLWSEQQVCQWLLWATNEFSLVNVNLHQFCMNGQMLCNLGKERFLELAPDFVGDLLW
                     EHLEQMIKENQEKTEDQYEENSHLNAVPQWINSNTLGFSMEQAPYGMQAPNYPKDNLL
                     DSMCPPSATPAALGSELQMLPKSRLNTVNVNYCSISQDFPSSNVNLLNNNSCKPKDHD
                     SPENGGDSFESSDSLLRSWNSQSSLLDVQKVPSFESFEEECSQSLCLSKLIMSEKDYI
                     QERSDPVEQGKPVIPAAVLAGFTGSGPIQLWQFLLELLSDESCQSFISWTGDGWEFKL
                     ADPDEVARRWGKRKNKPKMNYEKLSRGLRYYYDKNIIHKTSGKRYVYRFVCDLQNLLG
                     FTPEELHAILGVQPDTED"
BASE COUNT       847 a    913 c    847 g    889 t
ORIGIN


alignment_scores:
            Quality:  264.50             Length:    501
              Ratio:  1.260                Gaps:     19
 Percent Similarity:  41.916   Percent Identity: 22.954

alignment block:
 US-09-126-945B-2 x BC005486   ..

 Align seq 1/1  to: BC005486  from: 1  to: 3496

        8 LeuSerSerValSerProSerHisLeuLeuLeuProProAspThrValSe 24
          |||      ||||||||||  ||||||      |||             ||
       14 CTGAGGCCCGTCAGTCCCCGCCACCTTCCCCGGCCCCGCGCGCCCGGATC 63

       24 rArgThr.GlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArg 40
          |      ||||||      ::::::   |||   :::       ::||||
       64 GGCCCTACGGCCTC.....GTCTCGCCCGGCCTTGCGCGCCGGGACCGC 107

       41 ArgAsp...TrpSerPro.SerProProAlaThrProGlu.....GlnG 54
          ||||||      ||||||    ||||||||||   |||           :
      108 CGCGATCTCCTCTCCCCGCCGCCCTCCGGCTGGCCCTGCCTGCTGCGGCG 157

       54 lyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAsp 70
          ::::::::   |||   ::::::   :::|||::::::   |||
      158 CGATGAATGACTTTGGAATCAAGAACATGGACCAAGTGGCCCCCTGTCGCC 207

       71 SerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluProPr 87
          :::|||:::                |||:::   :::|||:::
      208 AACAGTTTTCGT...........GGCACACTCAAGCGCCAGCCAGCCTT 245

       87 oGlu......GluProGluGlnCysProValIleAspSerGlnAlaProA 102
            :::       :::   :::          |||::   |||   :::
      246 TGACACCTTCGATGGCTCTCTGTTTGCTGTGCTCCCTTCTCTCAGTGAAG 295

      102 laGlySerLeuAspLeuValPro:GlyGlyLeuThrLeuGluGluHisSer 118
            :::|||:::   ||||||   ||||||          :::|||:::
      296 ATCAGACACTCCAAGAAGTGCCCACGGGCCTGGATTCTGTCTCCCATGAC 345

      119 .....LeuGluGlnValGlnSerMetValVa 127
                        |||         ::::::   ||||:
      346 TCGGCCAGCTGCGAGCTGCCTTTGCTCACTCCCTGCAGCAAGGCAGTGAT 395

      127 lGlyGlnValLeuLys.....AspIleGlnThrAlaCysLysL 140
          ::::::::::||||||          ::::::         :::
      396 GAGCCAACCCTTAAAAGCCACCTTCAGTGGCTTCCAAAAGGAGCAACGAC 445

      140 euLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValGln 156
            |||:::|||      :::|||      ||||||   :::::::|||
      446 CTCTTGGCATCCCCAAAAACCCCTGGCTCTCGCAGCGAGCAGCAGGTGTGC 495

      157 LysTrpLeuLeuTrpThrGlnHisGlnTyrArgLeuProProMetGlyLy 173
          :::||||||||||||:::   :::::::::   |||      ::::::
      496 CAGTGGCTTCTCTGGGCCACCAACGAGTTCAGCCTGGTGAATGTGAACCT 545

      173 sAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGluGlnG 190
            :::   ||:::   ||||||   :::::::::|||:
      546 GCACCAGTTTGGCATGAACGGCCAGATGCTGTGTAACCTCGGCAAGGAGC 595

      190 lnPheArgGluArgSerPro...LeuGlyGlyAspValLeuHisAlaHis 205
          ::|||   :::   :::|||   :::   ||||||:::|||      |||
      596 GCTTCCTGGAGCTGGCGCCTGACTTTGTGGGTGACATCCTCTGGGAACAT 645

      206 LeuAsp..................................................... 207
          |||:::
      646 CTAGAGCAGATGATCAAAGAGAACCAAGAAAAGACAGAAGACCAATATGA 695

      208 .............IleTrpLysSerAlaAlaTrpMetLysGluArgThr. 219
                      :::                |||:::::::::::|||
      696 GGAAAACTCTCACCTCAACGCGGTTCCTCAGTGGATCAACAGCAATACAT 745

      219 ..................................................... 219

      746 TAGGCTTCAGCATGGAACAGGCTCCATATGGAATGCAGGCACCAAACTAC 795

      220 ..........................................SerProGl 222
                                                   :::|||:::
      796 CCCAAAGACAATCTCCTGGACAGCATGTGCCCGCCATCGGCCACGCCTGC 845

      222 yAla..................................................... 223
          ::::
      846 AGCTCTGGGCTCTGAGCTCCAGATGTTGCCCAAGTCTCGGCTCAACACCG 895

      224 .....IleHisTyrCysAlaSerThrSerGlu..................... 242
               :::::::|||||:::   :::::::::
      896 TCAATGTCAATTACTGTTCCATCAGCCAGGACTTCCCCAGCAGCAACGTG 945

      242 ..................................................... 242

      946 AATTTGCTCAACAACAATTCTGGAAAACCCAAGGACCACGACTCTCCAGA 995

      243 ..........................................GluSerTrpT 256
                                                   :::|||||||:
      996 GAACGGTGGGGACAGCTTCGAGAGCTCCGACTCGCTGCTGAGGTCCTGGA 1045

      256 hrAsp.....................................SerGlu 269
          :::::                                    ::::::
     1046 ACAGCCAGTCGTCCCTACTGGATGTACAGCGGGTACCTTCCTTCGAGAGC 1095
```

```
      249 ValAspSerSerCysSer..................................... 257
          :::::::::||||||
     1096 TTTGACGACGACTGTAGCCAGTCTCTGTGCCTCAGTAAGCTGACCATGTC 1145

      245 ..................................................... 245

     1146 CTTCAAGGACTACATCCAAGAGAGGAGCGACCCAGTCGAGCAAGGCAAAC 1195

      246 .....................................GlyGlnProIle 249
                                                 :::   ||||||
     1196 CAGTTATTCCTGCAGCAGTACTGGCTGGCTTCACTCGAAGCGGACCAATC 1245

      250 HisLeuTrpGlnPheLeuLysGlnGluLeuArgLysProSerTyr 293
          :::|||||||||||||||   |||   |||||||::      |||
     1246 CAGTTGTGGCAGTTTCTTCTGCAG..CTACTCTCTGACAAGTCCTGTCA 1292

      294 yArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysPheIleA 293
          ||||||   |||   :::   :::         |||||||::   |
     1293 ATCTTTCATCAGCTGGACGGGG..GATGGATGGGAGTTCAAGCTTGGTC 1339

      293 spSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAla 297
          ||      :::|||||||||   ||||||   |||||||||:::|||
     1340 ACCCCGATGAGCTTGCCCGCCGGTGGGGGAAGACGAAAAATAAACCAAAG 1389

      300 MetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGl 416
          ||||||||||:|||||||||||||:::::::   ||||||   ||||:
     1390 ATGAACTACGAGAAGCTGAGCCGGGGCTTACGTTACTACTACAAGAAGAA 1439

      316 yIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal 332
          :||||||:::|||   :::   :::::::|||   |||||:::|||||
     1440 GATCATCCACAAGACTTCG...GGCAAGCGCTACGTGTACCGTTTCGTA 1485

Seq_name: qb_rol:BC005504

Seq_documentation_block:
LOCUS       BC005504      3434 bp    mRNA            ROD       05-APR-2001
DEFINITION  Mus musculus, E26 avian leukemia oncogene 2, 3' domain, clone
            MGC:7483, mRNA, complete cds.
ACCESSION   BC005504
VERSION     BC005504.1  GI:13529586
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3434)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) med@pavil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at  http://image.llnl.gov
            Series: IRAK Plate: 8 Row: g Column: 7.
FEATURES             Location/Qualifiers
     source          1. .3434
                     /organism "Mus musculus"
                     /db_xref "taxon:10090"
                     /clone "MGC:7483 IMAGE:3490813"
                     /tissue_type "Mammary tumor, WAP IGF alpha model, 7 months
                     old, gross tissue."
                     /clone_lib:"NCI_CGAP_Mam5"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             100. .1506
                     /codon_start 1
                     /product "E26 avian leukemia oncogene 2, 3' domain"
                     /protein_id "AAH05504.1"
                     /db_xref="GI:13529587"
                     /translation="MGLRRVKMSVAGVANAELEKFRLQAHEIEKPFTVTVRTNRR
                     EDQTLQEVPTGLDSVSHDSASCELPLLDPCSKAVMSQALKAISSEGCRPSGRHSFN
                     PWIWSPDGVCQWLTWATNFPSLVNVNLHQFGMNFQMLCNLCKERFLELAPDFVGDTLW
                     EHLEQMTKENQEKTEDQYEENSHLNAVPPWTNSNTLGFSMEQAPYCMQAPNYPKDNIT
                     DGMCMTPSATPAAIGSEICMLPKSELNTVNVNYCSTSQPFTSGNVNKLCNNNSGKPFDRD
                     SPENGGDSFESSDSLTRSWNSQSSLLDVQRVPSFESEEFDCSQSLCLSKTIMSEKDYT
                     QFESFPVFLNQLEVTLPAAVLAGHTGSGRTQLWQFSLELLQSGKSCQSFISWTGQWEEKI
                     ADPDEVAPPWCKPPNEKEMNYEPLSRCLDYYYDENTIHKTSCPPYVYKFVCDTVNELS
                     FTPEELHAILGVQPDTED"
BASE COUNT       848 a    882 c    829 g    885 t
ORIGIN


alignment_scores:
          Quality:  264.00              Length:    486
            Ratio:  1.263                 Gaps:     18
 Percent Similarity:  43.004   Percent Identity: 22.222

alignment_block:
 US-09-126-945B-2 x BC005504    ..

 Align seq 1/1  to: BC005504  from: 1  to: 3434

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 86
         |||   :::   |||||   :::|||   :::|||::::::::::::|||::
       7 CCIACGGCCTCGACGCCGGCCTCG...CGCGCCGGCACGCCGCGAI 53

      86 IGlyLeuGluArgArgAspTrpSerProSerProProAlaThrProGlu. 52
         ::::      ||||||                |||||||||   |||
      54 CICCICICCCCGCCGC..............CCTCCGGCTGGCCCTGCCI 88

      53 .....GlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyr 67
               ::::::::::   |||   ::::::   :::|||::::::
      89 GCIGCGGCGCGAIGAATGACTTTGGAAICAAGAACATGGACCAAGIGGCC 138

      68 ProGluAspSerSerIrpAlaAlaLysAlaProGlyAlaSerSerArgGl 84
         |||      :::|||:::                |||:::   :::|||::
     139 CCIGTCGCCAACAGTTTTCGT.............GGGACACTCAAGCGCCA 176

      84 uGluProProGlu......GlnProGlnGlnCysProValIleAspSerG 99
         :      :::   :::   :::             |||:::   |||
     177 GCCAGCCTTGACACTTCGATGGCTCTCTGTTGCIGTGCTCCCTTCIC 226

      99 lnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrLeuGlu 119
           :::         :::|||:::   ||||||   ||||||
     227 TCAGTGAAGATCAGACACTCCAAGAAGTGCCCACGGGCCTGGATTCTGTC 276

     116 GluHisSer.....................LeuGluGlnValGlnSe 124
         :::|||:::                    |||         :::::
     277 TCCCATGACTCGGCCAGCTGCGACGTGCCTTTGCTCACTCCCTGCAGCAA 326

     124 rMetValValGlyGluValLeuLys.............AspIleGluThrA 137
         :   |||:::::::::::::|||||              ::::::
     327 GGCAGIGATGAGCCAAGCCTTAAAAGCCACCTTCAGTGGCTTCCAAAAGG 376

     137 IrCysLysLeuGlnAsnIleThrAlaAspProMetAspTrpSerProSer 153
           :::   |||:::|||       :::|||      ||||||   :::
     377 AGCAACGACGTCTTGGCATCCCCAAAAACCCCTGGCTGTGGAGCGAGCAG 426

     154 AsnValGlnLysTrpLeuLeuTrpThrGluHisGlnIyrArgLeuProPr 170
         :::|||   :::|||||||||||||:::   :::::::::   |||
```

```
 427 CAGGTGTGCCAGTGGCTTCTCTGGGCCACCAACGAGTTCAGCCTGGTGAA 476

 170 oMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMetS 187
      ::::::                  :::   |||:::   ||||||   ::::
 477 TGTGAACCTGCACCAGTTTGGCATGAACGGCCAGATGCTGTGTTAACCTCG 526

 187 crGluGluGlnPheArgGlnArgSerPro...LeuGlyGlyAspValLeu 202
     :::::|||:::|||   :::   :::|||   :::   ||||||:::|||
 527 GCAAGGAGCGCTTCCTGGAGCTGGCGCTGACTTTGTGGGTGACATCCTC 576

 203 HisAlaHisLeu.....................................As 207
           ||||||                                    ||
 577 TGGCAACATCTACAGCAGATGATCAAAGAGAACCAAGAAAAGACAGAAGA 626

 207 pIleTrpLysSerAlaAla.....................TrpMetLysG 217
     |   :::::::::   :::                    |||::::::::
 627 CCAATAGAGGAAAACTCTCACCTCAACGCGGTTCCTCATTGGATCAACA 676

 217 luArgThr..................................................... 219
     :::::|||
 677 GCAATACATTAGGCTTCAGCATGGAACAGGCTCCATATGGAATGCAGGCA 726

 219 ..... 219

 727 CCAAACTACCCCAAAGACAATCTCCTGGACAGCATGTGCCCGCCATCGGC 776

 220 .SerProGlyAla..... 223
      :::|||:::|||
 777 CACGCCTGCAGCTCTGGGCTCTGAGCTCCAGATGTTGCCCAAGTCTCGGC 826

 224 ..............IleHisTyrCysAlaSerThrSerGlu..... 232
                   ::::::||||||:::   :::::::::
 827 TCAACACGGTCAATGTCAATTACTGTTCCATCAGCCAGGACTTCCCCAGC 876

 242 ..... 242

 877 AGCAACGTGAATTTGCTCAACAACAATTCTGGAAAACCCAAGGACCACGA 926

 233 .....G 233
                                                      :
 927 CTCTCCAGAGAACGGTGGGGACAGCTTCGAGAGCTCCGACTCGCTGCTGA 976

 233 luSerTrpThrAsp..... 237
     ::|||||||:::::
 977 GGTCCTGGAACAGCCAGTCGTCCCTACTGGATGTACAGCGGGTACCTTCC 1026

 238 ...SerGluValAspSerSerCysSer..... 245
       ::::::   :::::::::|||||
1027 TTCGAGAGCTTTGAGGAGGACTGTAGCCAGTCTCTGTGCCTCAGTAAGCT 1076

 245 ..... 245

1077 GACCATGTCCTTCAAGGACTACATCCAAGAGAGGAGCGACCCAGTCGAGC 1126

 246 .....Gly 246
                                                      :::
1127 AAGGCAAACCAGTTATTCCTGCAGCAGTACTGGCTGGCTTCACTGGAAGC 1176

 247 GlnProIleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHi 264
       ||||||:::|||||||||||||||||   |||   |||||||:::
1177 GGACCAATCCAGTTGTGGCAGTTTCTTCTGGAG...CTACTCTCTGACAA 1223

 265 sSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheL 280
      |||        ||||||   |||   :::   :::        ||||
1224 GTCCTGTCAATCTTTCATCAGCTGGACGGGG...GATGGATGGAGTTCA 1270

 280 ysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsn 296
     ||:::   |||       :::|||||||||   ||||||   |||||||||
1271 AGCTTGCTGACCCCGATGAGGTTGCCCGCCGGTGGGGGAAGAGGAAAAAT 1320

 297 ArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTy 313
     :::|||   |||||||||:::||||||||||||||:::::|||   |||||
1321 AAACCAAAGATGAACTACGACAAGCTGAGCCGGGGCTTACGTTACTACTA 1370
```

```
 313 rLysLysGlyIleIleArgLysProAspIleSerGluArgLeuValTyrG 330
      |   |||:::|||||||:::|||   :::   :::::|||   ||||||:
1371 CGACAAGAACATCATCCACAAGACTTCG...GGCAAGCGCTACGTGTACC 1417

 330 lnPheVal 332
     ::|||||
1418 GTTTCGTA 1425

Seq_name: qb_pat2:189256

Seq documentation block:
LOCUS       189256        4240 bp    DNA              PAT       10-AUG-1998
DEFINITION  Sequence 4 from patent US 5721113.
ACCESSION   189256
VERSION     189256.1  GI:3409196
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4240)
  AUTHORS   Libermann,T.Aron, Oettgen,J.Peter and Kunsch,C.A.
  TITLE     NERF genes
  JOURNAL   Patent: US 5721113-A 3 24-FEB-1998;
FEATURES             Location/Qualifiers
     source          1..4240
                     /organism="unknown"
BASE COUNT     1070 a    592 c    662 g    916 t
ORIGIN


alignment_scores:
            Quality:  263.00              Length:     845
              Ratio:  1.584                 Gaps:      11
 Percent Similarity:  48.116   Percent Identity: 24.058

alignment_block:
 US-09-126-945B-2 x 189256   ..

 Align seq 1/1  to: 189256  from: 1  to: 4240

  20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa 36
     |||:::   ::::::   |||:::            ::::::|||:::::
 192 CCTCAGGGAATAACAATGACATCAGCAGTGGTTGACAGTGGAGGTACTAT 241

  36 l.................GlyLeuGluArgArgAspTrpSer..... 44
     :                  |||:::|||:::::::::   |||
 242 TTTGGAGCTTTCCAGCAATGGAGTAGAAAATCAAGAGGAAAGTGATAAGG 291

  45 ..............................ProSerProProAlaThrPro 51
                                    |||   |||   |||
 292 TTTCTGAATATCCAGCAGTGATTGTGGAGCCAGTTCCAAGTGCCAGATTA 341

  52 GluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyr.. 67
     |||||||||   :::|||      ||||:|||   |||      |||
 342 GAGCAGGGGTATGCAGCCCAGGTTCTGGGTTATGATGATGAGACTTATTT 391

  68 .ProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG 84
      :::|||   :::                                    |
 392 GATGCAAGATGTGGCA.................................G 408

  84 luGluProProGluGluProGluGlnCysProValIleAspSer..... 98
     |||||         |||   ||:::::::   :::::::::::::
 409 AAGAACAAGAAGTTGAGACCGAGAATGCGGAAACAGTGGAAGCATCACTT 458

  99 .....................................GlyAlaProAlaGl 104
                                          ::::||      ::
 459 CACAGCAGTAATGCACACTGTACAGATAAGACAATTGAAGCTGCTGAAGC 508

 103 ySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuG 120
     :   |||   :::                        ::::::::::||   |
 509 CCTGCTTCATATGGAATCTCCTACCTGCTTGAGGGATTCAAGAAGTCCTG 558
```

```
     120 lnGlnValGlnSerMetValValGlyGluValLeuLysAspIleGluThr 136
         ||   :::::::::   :::      :::||:::          ||||||
     559 AATTCATCCATGCTGCTATGAGGCCAGATGTCATT......ACAGAAACT 602

     137 AlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSe 153
         :::   ::::::   :::      :::::|||||||||   ||||||
     603 GTAGTGGAGGTGTCAACTGAAGAGTCTGAACCCATGGATACCTCTCCT.. 651

     154 rAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProP 170
                                                 ::::||
     652 .....ATTCCAA 657

     170 roMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaMet 186
                 :::   :::   :::      ||:::          :::
     658 CATCACCAGATAGCCATGAACCAATGAAAAAGAAAAAA.....GTT 698

     187 SerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuHi 203
         :::::::::         |||:::||||||:::::::::
     699 GGCCGTAAACCAAAGACCCAGCAATCACCAATTTCCAATGGG..... 740

     204 sAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrS 220
                       :::          :::|||:::::
     741 .................TCTCCTGAGTTAGGTATAAAGAAGAAACCAA 771

     221 erProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpIle 236
           |||
     772 GAGAAGGA..... 779

     237 AspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpCl 253
                           :::|||:::      :::||||||::
     780 .....AAAGGAAACACAACCTATTTGTGGGA 805

     254 nPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleA 270
         ::|||||   :::|||||          ::::::          ||||:::|||:
     806 GTTTCTTTTAGATCTACTTCAAGATAAAAATACTTGTCCCAGGTATATTA 855

     270 rgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGln 286
         ::|||   :::::|||||||||||||||||||:::   ||||||
     856 AATGGACTCAGAGAGAAAAAGGCATATTCAAGCTGGTGGATTCAAAGGCT 905

     287 ValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAs 303
         |||:::::|||||||||   :::||||||:::|||   ||||||||||:
     906 GTCTCTAACTTTGGGGAAAGCATAACAACAAACCAGACATGAACTATGA 955

     304 pLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgL 320
         :   ::::::|||:::::::|||   ||||||:::::::|||||:::   |
     956 AACCATGGCACGAGCTTTGAGATACTACTACCAAAGGGCAATTCTTGCAA 1005

     320 ysProAspIleSerGlnArgLeuValTyrGlnPhe 331
         ||   :::   :::|||||||||||||||||||||
    1006 AGCTTGAA   GGACAGAGGCTTGTATATCAGTTC 1037

seq_name: qb_pat1:AR103482

seq_documentation_block:
LOCUS       AR103482     5427 bp    DNA              PAT       14-FEB-2001
DEFINITION  Sequence 2 from patent US 6087485.
ACCESSION   AR103482
VERSION     AR103482.1  GI:12815070
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5427)
  AUTHORS   Brooks Wilson,A.R., Buckler,A., Cardon,L., Carey,A.H., Galvin,M.,
            Miller,A. and North,M.
  TITLE     Asthma related genes
  JOURNAL   Patent: US 6087485-A 2 11-JUL-2000;
FEATURES             Location/Qualifiers
     source          1..5427
                     /organism "unknown"
BASE COUNT     1617 a   1047 c   1147 g   1616 t
ORIGIN
```

```
alignment_scores:
            Quality:  262.00              Length:     868
              Ratio:  1.424                 Gaps:      18
 Percent Similarity:  50.689   Percent Identity:  25.895

alignment_block:
 US-09-126-945B-2 x AR103482   ..

 Align seq 1/1  to: AR103482  from: 1  to: 5427

      41 ArgAspTrpSerProSerProProAlaThrProGluGlnGly..... 54
         |||  |||  |||  |||  ::::::  |||  :::
      88 CGCCCGTGGTGCCCCATCCCTATAGGAGCTGGTGAGATTGCAGCCTGCTG 87

      55 .LeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGlu.... 69
          |||  ::::::      |||||||::  :::  :::|||||
      88 CCTCCCCTCCATCAGCCACAGCTATTGG...ATTTCCCACCCAGAATCTT 134

      70 ................AspSerSer.....TrpAlaAlaLysAlaPro 78
                        |||||||::  ||||::  :::
     135 TAGGTAAATGAGATCATGATTCTGGAAGGAGGTGGTGTAATGAA..... 178

      79 GlyAlaSerSerArgGluGluProProGluGluProGluGlnCysProVa 95
         ::::::  |||:::::::|||||
     179 ...TCTCAACCCCGGCCAACAACCTCCT..... 202

      96 lIleAspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyL 112
                 :::|||||||::|||||||||
     203 .........TCACCAGCCGCCAGCCTGGAC..... 223

     113 euThrLeuGluGlu.HisSerLeuGluGlnValGlnSerMetValValGl 128
             :::::: :::|||     :::|||:::|||          ||
     224 .......AGACAGCTACTCCACGTGCAATGTTTCCAGTGGGTTTTTTGG 265

     129 yGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrA 145
         |           :::|||:::
     266 AGGCCAGTGGCATCAAATTCAT..... 287

     146 laAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrp 161
             |||:::   |||:::      :::|||   :::|||||||
     288 .....CCTCAGTACTGGACCAACTACCAGGTGTGGGACTGGCTCCAGCAC 332

     162 ThrGluHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu.. 177
                     :::|||                |||||||||
     333 CTCCTGGACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTT 382

     178 ....LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgG 193
             :::  |||:::::::|||||:::|||||   :::::::||   :
     383 CGACATCAACGGCGAGCACCTCTGCAGCATGAGTTTGCAGGAGTTCACCC 432

     194 InArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu..... 209
         ::   :::      :::|||:::::::|||:::::::|||
     433 GGGCGGCAGGGACGGCGGGGCAGCTCCTCTACAGCAACTTGCAGCATCTG 482

     210 .....AspIleTrpLysSer.....Al 212
                              |||:::::::::|||          ::
     483 AAGTGGAACGGCCAGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGT 532

     213 aAlaTrpMetLysGluArgThrSerProGlyAlaIle..... 224
         :          |||:::|||:::|||:::   :::
     533 CATTGTCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAG 582

     225 .....HisTyrCysAlaSerThrSerGluGluSerTrpIleAspSerGlu 239
          :::|||      :::|||:::      |||  :::|||
     583 ACGAGAACTATTATATGACACCAACTATGGTAGCACAGTAGATTTG... 629

     240 ValAspSerSer.....CysSerGlyGln..... 247
         :::|||||:::      |||   :::|||
     630 TTGGACAGCAAAACTTTCTGCCGGCCTCAGATCTCCATGACAACCACCAG 679

     248 .......ProIle..... 249
```

```
                ||||::
      680 TCACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCC 729

      250 ..........................................HisLeuTrp 252
                                                   |||||||||
      730 CTGCCAAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGG 779

      253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl 269
          :::|||::::::::::::::|||||||::|||   :::          :::||
      780 GAATTCATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAAT 829

      269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaG 286
          |:::|||   :::::::::::::|||:::|||::::::      |||
      830 AAAATGGCAAGACCGATCTGAGGCGTCTTCAGGTCTTGAAATCAGAGG 879

      286 InValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
            ||||||||:|||||||||   :::|||||||||:   :::|||||:|||
      880 CAGTGGCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTAT 929

      303 AspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 319
          :::|||||||||||||||:::|||   ||||||||||::   |||::::::
      930 GAAAAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATACTGGA 979

      319 qLysProAspIleSerGlnArgLeuValTyrGlnPhe 341
          ::::   |||   ::::::|||||||||||:|||
      980 GCGTGTGGAT...GGACGAAGACTGGTATATAAATTT 1013

seq_name: qb_pat1:AR103483

seq_documentation_block:
LOCUS       AR103483     5510 bp    DNA              PAT       14 FEB 2001
DEFINITION  Sequence 3 from patent US 6087485
ACCESSION   AR103483
VERSION     AR103483.1  GI:12815071
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5510)
  AUTHORS   Brooks Wilson,A.R., Buckler,A., Cardon,L., Carey,A.H., Galvin,M.,
            Miller,A. and North,M.
  TITLE     Asthma related genes
  JOURNAL   Patent: US 6087485 A 3 11 JUL 2000;
FEATURES             Location/Qualifiers
     source          1..5510
                     /organism="unknown"
BASE COUNT     1624 a    1064 c    1191 g    1631 t
ORIGIN


alignment scores:
            Quality:  262.00            Length:     329
              Ratio:  1.497               Gaps:      13
 Percent Similarity:  53.191   Percent Identity: 25.228

alignment block:
 US-09-126-945B-2 x AR103483  ..

 Align seq 1/1  to: AR103483  from: 1  to: 5510

       71 SerSerTrpAlaAlaLysAlaProGlyAlaSerSer......ArgGluGl 85
          |||:::|||::::::::::::   |||::::::          :::   ||
      152 TCAGCTTGGGTAAGCCGGGTCTGCGGCGGGCATTGCCATCTGAAGACAGA 201

       85 nProProGluGluProGluGlnCysProValIleAspSerGlnAlaProA 102
          |         :::|||:::   :::   ::::::       ::::::   :
      202 GGCAGGAGGGCAGCCACACCTTGCCCAGATCATGATTCTGGAAGGA...G 248

      102 laGlySerLeuAspLeuValProGlyGlyLeuThrLeu............ 114
          ::|||   ::::::::|||   ||||||||:::      |||
      249 GTGGTGTAATGAATCTCAACCCCGGCAACAACCTCCTTCACCAGCCGCCA 298

      115 ........GluGluHisSerLeuGluGlnValGlnSerMetValValGl 128
                  :::::::::|||      :::|||:::|||          ||
      299 GCCTGGACAGACAGCTACTCCACGTGCAATGTTTCCAGTGGGTTTTTGG 348

      129 yGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrA 145
          |           :::::::::
      349 AGGCCAGTGGCATGAAATTCAT.......................... 370

      145 laAspProMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrp 161
            |||:::   |||:::      :::|||   ::::::::
      371 .....CCTCAGTACTGGACCAAGTACCAGGTGTGGGAGTGGCTCCAGCAC 415

      162 ThrGlnHisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu.. 177
                      :::|||                |||||||||
      416 CTCCTGGACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTT 465

      178 ....LeuAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgG 193
          ::::   |||::::::|||||||:||||||   :::::::||   :
      466 CGACATCAACGGCGAGCACCTCTGCAGCATGAGTTTGCAGGAGTTCACCC 515

      193 InArgSerProLeuGlyGlyAspValLeuHisAlaHisLeu......... 206
          ::   :::      :::|||::::::|||:::::::::|||
      516 GGGCGGCAGGGACGGCGGGCAGCTCCTCTACAGCAACTTGCAGCATCTG 565

      207 .........................AspIleTrpLysSer.........AI 212
                                    |||:::::::::|||        ::
      566 AAGTGGAACGGCCAGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGT 615

      212 aAlaTrpMetLysGluArgThrSerProGlyAlaIle.............. 224
          :       |||:::|||::::::::   :::
      616 CATTGTCAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAG 665

      225 .....HisTyrCysAlaSerThrSerGluGluSerTrpThrAspSerGlu 239
           :::|||      :::||||::      |||   :::|||
      666 ACGAGAACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG... 712

      240 ValAspSerSer......CysSerGlyGln..................... 247
          :::||||||::      |||   :::|||
      713 TTGGACAGCAAAACTTTCTGCCGGGCTCAGATGTCCATGACAACCACCAG 762

      248 .......ProIle..................................... 249
                |||:::
      763 TCACCTTCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCC 812

      250 ..........................................HisLeuTrp 252
                                                   |||||||||
      813 CTGCCAAGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGG 862

      253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl 269
          :::|||::::::::::::::|||||||::|||   :::          :::||
      863 GAATTCATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAAT 912

      269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaG 286
          |:::|||   :::::::::::::|||:::|||::::::      |||
      913 AAAATGGCAAGACCGATCTGAGGCGTCTTCAGGTCTTGAAATCAGAGG 962

      286 InValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
            ||||||||:|||||||||   :::|||||||||:   :::|||||:|||
      963 CAGTGGCTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTAT 1012

      303 AspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 319
          :::|||||||||||||||:::|||   ||||||||||::   |||::::::
     1013 GAAAAGCTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATACTGGA 1062

      319 qLysProAspIleSerGlnArgLeuValTyrGlnPhe 341
          ::::   |||   ::::::|||||||||||:|||
     1063 GCGTGTGGAT...GGACGAAGACTGGTATATAAATTT 1096

seq_name: qb_in3:DROE74B

seq_documentation_block:
LOCUS       DROE74B      5315 bp    mRNA             INV       26-APR-1993
DEFINITION  Drosophila melanogaster E74B protein.
ACCESSION   M37083
```

```
VERSION     M37083.1  GI:157309
KEYWORDS    E74B protein.
SOURCE      D.melanogaster (strain Canton S), cDNA to mRNA, mRNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 5315)
  AUTHORS   Burtis,K.C., Thummel,C.S., Jones,C.W., Karim,F.D. and Hogness,D.S.
  TITLE     The Drosophila 74EF early puff contains E74, a complex
            ecdysone-inducible gene that encodes two ets-related proteins
  JOURNAL   Cell 61, 85-99 (1990)
  MEDLINE   90199900
COMMENT     Draft entry and computer readable sequence for [1] kindly submitted
            by C.S.Thummel, 10 SEP-1990.
FEATURES             Location/Qualifiers
     source          1. .5315
                     /organism="Drosophila melanogaster"
                     /strain="Canton S"
                     /db_xref "taxon:7227"
                     /map "3L 74EF"
     source          1. .267
                     /organism "Drosophila melanogaster"
                     /strain "Canton S"
                     /db_xref "taxon:7227"
     gene            1052. .5304
                     /gene "E74B"
     CDS             1052. .3703
                     /gene "E74B"
                     /codon_start=1
                     /product "E74B protein"
                     /protein_id "AAA28494.1"
                     /db_xref "GI:157310"
                     /translation "MIMVQHIVAASAHNPASQAAASIVNVSSSSSSSSSSSSSSISLS
                     SSSSSSSLSSATPTPVASPVTPTSPPPAAAAPAKASPPAGAELQEDGQQQAKTQEDPTM
                     KDQDMLEKIKQEVKDPVNVEEPGAIVDTESVMAKQSPSPVASTEVPESLEEISNKSPP
                     VQEDEEESPSVASDCPEFKVIYNHIPQQQHHHSPSSPDKIRSTIIQVSKTIWPPKQQL
                     QRSSVITAAPTLQPQQHQQPMSDIEDEETLEDVDIAIADVEADAELEELLEQYQNGYD
                     SPLDLSLGGATSAAASAAAAASAVSKPPKGRTYSGTESDQSAQCEKAKMKLKPEPKAEP
                     SAAYKKSLMKRYYTEIPIVKQSTSPAPQQQLQQQQHHLQQQQQQQQPHNGSTPAGATALL
                     HIKTEQNILLETPLQLQQQQQQQQGIHGAAGNGGSSNGNNAHQQQQPLAIPQRPLLHNL
                     LSGGAIHNPHHRNYITATTGSFPPSPADSGVSDVDSSSSGGQPCADEIKARLGMPPAT
                     SASAAAAAAAAAAAAAHLHTGTFLHPNLYQNNAANGLRNIWNKSVGVPLNYYGSSGAG
                     SGGIQPGGPGNPQTPGYITTSYFNAPTAATAAASQRGTPINGYHSIHQQQQQQQQQSQQ
                     SQQQQQIAHQQISHQQQQAIHQQISHQQQQQQQQQQQHPHSQLNGPHPHSHPHSHPHS
                     HPHAGQRTHSTIAAAAAAAAASVVSSSSSAVAAAAMQSASAAAAAIAAAAAGGSQSVI
                     QFATSSVSYDLSYMIFLGGPQQPRAKPPKYPIPCVPKWSWQCSIIYLWKEIKPQ
                     DPEYQPPFIPWINKEKGVFKLVGSKAVSRLWGMHKNSIDMNYETMGRAILKYYQKGIL
                     AKVDGQRLVYQFVDVPKDIIELDCNGV"
     polyA_signal    5299. .5304
                     /gene "E74B"
BASE COUNT      1465 a    1423 c    1304 g    1123 t
ORIGIN


alignment_scores:
            Quality:  261.00              Length:     385
              Ratio:  1.305                 Gaps:      19
 Percent Similarity:  51.948   Percent Identity: 28.052

alignment_block:
 US-09-126-945B-2 x DROE74B  ..

 Align seq 1/1  to: DROE74B  from: 1  to: 5315

       2 GlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLe 18
         ||||||||||||||||  :::  ::||||  ||||||||||
    2614 GGCAGCGGCACACCTGCACACGGGCACCTTTCTGCATCCGAATTTATA.. 2661

      18 uProProAspThrValSerArgThrGly.....LeuGluLysAlaA 42
           ||:::      ::::::  :::         ::||||||  :
    2662 ...CCAGAACAATGCGGCAAATTCGCTGCGGAACATATGGAATCGCAGI 2707

      42 laAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSer..... 46
         ::  ||||||      |||  ||||||||||      :::
    2708 GTTCGCCTCGCCGCACAACTACTACCCAACGCACTCGCCGCCGCCGCAGTCGCCGC 2757

      47 ...Pro.ProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeu.... 60
           ||| ||:::  |||  :::     :::      ||||||
    2758 CACCCAGCCAGGTCGTCCGGCAAACCCCCAAACCCCTGCCTACCTGACAA 2807

      61 ..SerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLys 76
         ||||||||||:::
    2808 CGTCCTACTTCAAC.....................................  2821

      77 AlaProGlyAlaSerSer.ArgGluGluProProGluGluProGluGlnC 93
         ||||||   |||:::::: ::::::::::|||   ||||||||||:::::
    2822 GCACCGACAGCGGCAACGCCAGCGCCTCCCCACCGACGAACCACAATCAA 2871

      93 ysProValIle.AspSerGlnAlaProAlaGlySerLeuAspLeuValPr 109
           :::|||   :::   |||   |||||:::::      ::::::
    2872 TGGCTATCATCCCTGCACCAGCAGCAGCAGCAGCAACAGCAGTCGCAGC 2921

     109 oGlyGlyLeuThrLeuGluGluHisSerLeuGluGluValGlnSerMetV 126
           :::   :::                          |||        :
    2922 AATCGCAGCAGCAACAGCAACTGGCGCACCAGCAACTGTCGCACCAGCAG 2971

     126 alValGlyGlnValLeuLysAspIleGluIleAlaCysLysLeuLeuAsn 142
         ::::::|||   :::         :::      ::::::        :::
    2972 CAGCAGGCGTTGCATCAGCAACTGTCGCACCAACAGCAGCAGCAGCAACA 3021

     143 IleThrAlaAsp.............ProMetAspTrpSerProSerAsnVa 155
           ||:::|                 |||  :::   |||  |||  ::
    3022 GCAACAGCAGCAACACCGCCACTCCCAACTGAACGGTCCGCATCCGCACT 3071

     155 lGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetG 172
         :  :::  |||      :::     :::  |||  |||      :
    3072 CACATCCGCACTCTCATCCGCATTCGCATCCGCACGCGGCGCCA.....G 3115

     172 lyLysAlaPheGlnGlnGln              AlaGlyLysGluLeuCys 184
         ::  ||||||:::::::                 :::|||||||||||||
    3116 CACACGCATTCCACAATCGCTGCAGCGGGGGCAGCGGCAGCAGCATCTGT 3165

     185 AlaMetSerGluGluGlnPheArg.................GlnArgSe 195
         :::::::::::::|||   |||                  :::|||::
    3166 CGTCAGCAGCAGCAGCAGCGCCGTTGCAGCGCACGCATCCTGCAGCGCA 3215

     195 rProLeuGly..................................... 198
         :   |||
    3216 CTCCAGCGCCAGCAGCAACCCCGCCACGTCGCGGGGGGCGCCATCCCACAGT 3265

     199 ...........GlyAspVal.....................LeuHisAla 204
                    |||||||||                  |||||||||
    3266 GTTATCCAACCGGCGACGTCCAGCGTCAGCTACGATCTCTCCTACATGCT 3315

     205 HisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerPr 221

    3315 ..................................................  3315

     221 olyAlaIleHisTyrCysAlaSerThrSerGluGlu...SerTrpThrA 237
          ||||.|   :::      ||||::  ::||||||   :::   |||:
    3316 GGAGCTGGCGCGCATTCCAGCAGCGCAAGCGGAAGAAGCCGCGCAAACCG 3364

     237 spSer.GluValAsp.........SerSerCysSerGlyGlnProIleHi 250
         ::::: |||::                |||  ::|||:::      ::
    3365 AAGCTGGAGATGGGCGTAAAGCGGCGCAGCCGGGAGGGATCCACCACCTA 3414

     250 sLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyA 267
         :|||||:|::|||||   :::|||||         :::         |
    3415 TCTGTGGGAGTTCCTCCTCAAACTGCTCCTGAAGATCGCGAATACTGTCCGC 3464

     267 rgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAsp 283
         ||||||||||:::|||   |||:::||||||||||:::||||:::   |||
    3465 GTTTCATCAACTGGACCAACCGGGACAAGGGCGTCTTCAAGCTGGTCGAC 3514

     284 SerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMe 300
```

```
       |||      |||:::|||||||||||||:::::|||||||:::|||   ||
    3515 TCGAAGGCTGTGTCCCGCCTGTGGGGCATGCACAAGAACAAGCCGGACAT 3564

     300 tAsnTyrAspLysLeuSerArgSerIleArgGlaTyrTyrLystLysGlyI 317
         |||||||:::   :::::|||:::::|||   |||||:::::|||
    3565 GAACTACGACGACGATGGGCAGGGCGCTGAGGTACTACTACCAGCGCGGCA 3614

     317 IeIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal 332
         ||:::   |||   |||   :::|||||||||||||||||||||||
    3615 TACTGGCCAAGGTGGAT...GGCCAGCGGCTGCTCTACCAGTTCGTG 3658

seq_name: qb_in4:DROE74A

seq_documentation block:
LOCUS       DROE74A       6479 bp    mRNA            INV       26-APR 1993
DEFINITION  Drosophila melanogaster E74A protein.
ACCESSION   M37082
VERSION     M37082.1  GI:157307
KEYWORDS    DNA binding protein; E74A protein.
SOURCE      D.melanogaster (strain Canton S), cDNA to mRNA, and DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 6479)
  AUTHORS   Burtis,K.C., Thummel,C.S., Jones,C.W., Karim,F.D. and Hogness,D.S.
  TITLE     The Drosophila 74EF early puff contains E74, a complex
            ecdysone inducible gene that encodes two ets-related proteins
  JOURNAL   Cell 61, 85-99 (1990)
  MEDLINE   90199900
COMMENT     Draft entry and computer readable sequence for [1] kindly submitted
            by C.S.Thummel, 10-SEP 1990.
FEATURES             Location/Qualifiers
     source          1. .6479
                     /organism "Drosophila melanogaster"
                     /strain "Canton S"
                     /db_xref "taxon:7227"
                     /map "3L 74EF"
     gene            2378. .6468
                     /gene "E74A"
     CDS             2378. .4867
                     /gene "E74A"
                     /codon_start 1
                     /product "E74A protein"
                     /protein_id "AAA28493.1"
                     /db_xref "GI:157308"
                     /translation "MPFIDDALLWCPDNEGRLVGGLDLGTGIADDSTANGTENLNPSI
                     QSAGNPNNPQQSVGGETLGSVESAGNELNGAAARNVNVVVEPLCGGDSSDELFRSFSE
                     SNFETESLLSDLATVEVEKVENEENNNNVITTDEDFASVAAAVVANGDLLACENAQLSAG
                     GLVDSVAASLADSGDAGGQQALLAFGSSSSAASATAAAAAALGGDLLNNNNNNSNSNN
                     NSNGNGNHGGGGGASSGGGVAGDGATKLEYALMGGQPLAEEPREVTSAAANPLLVEK
                     LMSKCLNIEKRMDKLSDTETPIVKQSTSPAPQQQLQQQHHLQQQQQQQQPHNGSTFAGA
                     TALLHIKTFEQNTLLTPLQLQQQQQQQQGLHGAAGNGGSSNGNNAHQQQQPLAIDQRPL
                     LHNLLSGGATHNPHHRNYTTATTGSFPPSPADSGVSDVDSSSSGGQPCADELKARLGM
                     PPATSASAAAAAAAAAAAAAHLHTGTFLHPNLYQNNAANSLRNIWNRGVGVPPNYYGS
                     SGAGSGGTQPGGPGNPQTPGYLFTSYFNAPTAATAAASQRGTTINGYHSLHQQQQQQQ
                     QSQQSQQQQQLAHQQLSHQQQQAIHQQLSHQQQQQQQQQQHPHSQINGPHPHSHPHS
                     HPHSHPHAGQHTHSTIAAAAAAAAASVVSSSSSSAVAAAAMLSASAAAAATAAAAAGGS
                     QSVIQPATSSVSYDLSYMELEGGPQQRKAKKPRKPKLEMGVKQKSREGSTTYLWEFLL
                     KLLADREYCPRFIKWINREKGVFKLVDSKAVSKLWGMHKNKPQMNYETMGRALRYYYQ
                     RGILAKVDGQRLVYQFVDVPKDIIELDCNGV"
     polyA_signal    6464. .6468
                     /gene "E74A"
BASE COUNT     1879 a    1641 c    1481 g    1478 t
ORIGIN


alignment scores:
            Quality:  261.00              Length:     485
              Ratio:  1.405                 Gaps:      19
 Percent Similarity:  51.948   Percent Identity: 28.052

alignment block:
```

```
US-09-126-945B-2 x DROE74A   ..

Align seq 1/1  to: DROE74A  from: 1  to: 6479

      2 GlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLe 18
        ||||||:::::|||:::   :::   :::|||   |||:::::::::::
   3778 GGCAGCGGCACACCTGCACACGGGCACCTTTCTGCATCCGAATTTATA.. 3825

     18 uProProAspThrValSerArgThrGly.....LeuGluLysAlaA 32
          |||:::      ::::::   :::       :::|||:::   :
   3826 ....CCACAACAATGCGGCAAATTCGCTGCGGAACATATGGAATCGCAGT 3871

     32 laAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSer..... 46
        ::   ||||||      |||   |||:::::|||      :::
   3872 GTTGGGGTGCCCGACAACTACTACGGAAGCAGTGGCGCCGGCAGTGGGGC 3921

     47 ...Pro.ProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeu.... 60
           ||| |||:::   |||   :::      :::       ||||||
   3922 CACCCAGCCAGGTGGTCCGGGAAACCCCCAAACGCCTGGCTACCTGACAA 3971

     61 ..SerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLys 76
          ||||||||||:::
   3972 CGTCCTACTTCAAC..................................... 3985

     77 AlaProGlyAlaSerSer.ArgGluGluProProGluGlnProGluGluC 93
        ||||||   |||:::::: ::::::::::|||   ||||||||||:::::
   3986 GCACCGACAGCGGCAACGGCAGCGGCGTCCCAGCGAGGAACCACAATCAA 4035

     93 ysProValIle.AspSerGlnAlaProAlaGlySerLeuAspLeuValPr 109
          :::|||   :::   |||   |||::::::      ::::::
   4036 TGGCTATCATTCCCTGCACCAGCAGCAGCAGCAGCAACAGCAGTCGCAGC 4085

    109 oGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetV 126
          :::   :::                  |||          :
   4086 AATCGCAGCAGCAACAGCAACTGGCCCACCAGCAACTGTCGCACCAGCAG 4135

    126 alValGlyGlnValLeuLysAspIleGluThrAlaCysLysLeuLeuAsn 142
        :::::|||   :::       :::      ::::::          :::
   4136 CAGCAGGCGTTGCATCAGCAACTGTCGCACCAACAGCAGCAGCAGCAACA 4185

    143 IleThrAlaAsp.............ProMetAspTrpSerProSerAsnVa 155
          ||||||                |||   :::   |||   |||   ::
   4186 GCAACAGCAGCAACACCCGCACTGCCAACTGAACGGTCCGCATCCGCACT 4235

    155 lGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetG 172
        :   :::   |||       :::      :::   |||   |||    :
   4236 CACATCCCCACTCTCATCCGCATTCGCATCCCCACCGCGGCCA.....C 4279

    172 lyLysAlaPheGlnGluLeu.............AlaGlyLysGluLeuCys 184
        ::   ||||||:::::                 :::|||:::::::::|||
   4280 CACACGCATTCCACAATCGCTGCAGCGGCGGCAGCGGCAGCAGCATCTGT 4329

    185 AlaMetSerGluGluGlnPheArq.................GlnArgSe 195
          :::::::::::::|||   |||                 :::||||::
   4330 CGTCAGCAGCAGCAGCAGCGGCGTTGCAGCGGCAGCCATGCTGAGCGCCA 4379

    195 rProLeuGly..................................... 198
        :      |||
   4380 GTGCAGCGGCAGCAGCAACGGCGGCAGCTGCGGCGGCGGCGGATCCCAGAGT 4429

    199 ..........GlyAspVal.....................LeuHisAla 204
                |||||||||                     |||||||||
   4430 GTTATCCAACCGGCGACGTCCAGCGTCAGCTACGATCTCTCCTACATGCT 4479

    205 HisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrSerPr 221

   4479 ..................................................

    221 oGlyAlaIleHisTyrCysAlaSerThrSerGluGlu...SerTrpThrA 237
        ||||||   :::      ||||::   :::||||||   :::   |||:
   4480 GGAGCTGGGCGGATTCCAGCAGCGGAAGGCGAAGAAGCCGCGCAAACCG 4528

    237 spSer.GluValAsp.........SerSerCysSerGlyGlnProIleHi 250
```

```
      ::::: ||:::                ||| ::::|:::      ::
  4529 AAGCTGCAGATGGGCCTAAAGCGGGCGCAGCCGGGAGGGATCCACCACCTA 4578

   250 sLeuTrpGlnPheLeuLysGluLeuLeuLysProHisSerTyrGlyA 297
       ||||||:::|||||  :::|||||||          :::        |
  4579 TCTGTGGGAGTTCCTCCTCAAACTGCTCCAGGATCGCGAATACTGTCCGC 4628

   267 rqPheIleArqTrpLeuAsnLysGluLysGlyIlePheLysIleGluAsp 283
       |||||||||:::|||  |||:::||||||||||:::|||||:::  |||
  4629 GTTTCATCAAGTGGACGAACCGGGAGAAGGGCGTCTTCAAGCTGGTGGAC 4678

   284 SerAlaGlnValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMe 300
       |||      |||:::|||||||||||||:::::|||||||:::|||  ||
  4679 TCGAAGGCTGTGTCCCGCCTGTGGGGCATGCACAAGAACAAGCGGGACAT 4728

   300 tAsnTyrAspLysLeuSerArqSerIleArqGlnTyrTyrLysLysGlyI 317
       ||||||||:::  :::::|||:::::|||  |||||:::::|||||
  4729 GAACTACGACAAGCTGAGCAGGAGCATCAGGCAGTACTACAAGAAGGGCA 4778

   317 IelleArqLysProAspIleSerGlnArqLeuValTyrGlnPheVal 332
       |||::  |||  |||  :::|||||||||||||||||||||||||
  4779 TACTGGCAAGGTGGAT  GGCCACGGGCTGGTCTACCAGTTCGTG 4822

seq_name: qb_ov:GGCETS

seq_documentation_block:
LOCUS       GGCETS       1762 bp    mRNA            VRT       12-SEP-1993
DEFINITION  Chicken c-ets mRNA for p54 (c-ets) protein.
ACCESSION   X13026
VERSION     X13026.1  GI:63179
KEYWORDS    c-ets oncogene; cellular oncogene.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1762)
  AUTHORS   Dutertre-Coquillaud,M., Leprince,D., Flourens,A., Henry,C.,
            Ghysdael,J., Debuire,B. and Stehelin,D.
  TITLE     Cloning and expression of chicken p54c-ets cDNAs, the first
            p54c-ets coding exon is located into the 40.0 Kbp genomic domain
            unrelated to v-ets
  JOURNAL   Oncogene Res. 2 (4), 335-344 (1988)
  MEDLINE   88289026
COMMENT     See <x00144> for v-ets sequence from AEV E26 fragment.
FEATURES             Location/Qualifiers
     source          1. .1762
                     /organism "Gallus gallus"
                     /db_xref "taxon:9031"
                     /cell_type "spleen cell"
                     /clone_lib="lambda gt11"
                     /clone="pCK E54A"
     CDS             82. .1407
                     /note="p54 (c-ets) protein (AA 1-441)"
                     /codon_start 1
                     /protein_id="CAA31441.1"
                     /db_xref "GI:63180"
                     /db_xref "SWISS-PROT:P13474"
                     /translation "MKAAVDLEPTLIFIKEEVPDPEPSPEMPAAEVFLLFSSEEM
                     MSQALKATFSGFAKEQGPEGERKDPQQWIETHVPDWVMWAVNEFSLKGVDFQSFCMNC
                     AALCALGKECFLELKPDFVGDILWEHELETLQKEEAKPTTANGVNAAYFESKYTLDYFI
                     SYGIEHAQCVEPSFFSFESPLTFESVQTTHPLSSFFTISLKYFNLYESVTLKDDVQTIS
                     IQTDYFTIKQFVVTPDNMCMGKASKCKLGGQDSFESTESYDSCDKLTQSWSSQSSFQS
                     LQFVPSYDSFDSEDYPAALPNHKPKGIFKDYVPLFAADMNRDKPVTPAAALAGYTGSGF
                     IQLWQFTTELTDPSVQGTLSWTGSDWLFKLSTELVAEWLLKNGYEDRLMNLLSRG
                     LPYYYDKNIIHKTAGKPYVYDFEVCDPQSLLGYTPEELHAMLPVRPDADE"
BASE COUNT      444 a     479 c     465 g     374 t
ORIGIN


alignment_scores:
            Quality:  260.50            Length:    448
              Ratio:  1.480               Gaps:     17
```

```
 Percent Similarity:  49.286   Percent Identity: 24.214

alignment_block:
 US-09-126-945B-2 x GGCETS  ..

 Align seq 1/1  to: GGCETS  from: 1  to: 1762

       74 AlaAlaLysAlaProGlyAlaSerSerArqGluGluProProGluGlu.P 90
          ||||||  ||||||||:                :::::|||||   ::: |
       12 GCCGCCGCTGCCCCGGCG.............CAGCAGCCGCCCCCAGTTC 49

       90 roGluGlnLysProValIleAspSerGluAlaProAlaGlySerLeuAsp 106
          ||:::      ||||::      :::::::::      ::::::::|||
       50 CCAGTCCCGGTCCCTCGCTCCGGCCTCAACCATCAAGGGCGGTGGAC 99

      107 LeuValPro..................................... 109
          |||  |||
      100 CTGAAGCCCACCCTCAGCAICATCAAGACGGAGAAGGTCGACATCGATCT 149

      110 .....................GlyGlyLeuThrLeuGluGluHisS 118
                              :::  :::  |||        |
      150 CTTCCCGTCCCCCGATATGCAATCTGCAGATGTGCCTTTGTTAACCCCCA 199

      118 erLeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspIle 134
          ||  ::::::::::::|||:::::::::  :::::::::  :::
      200 GCAGCAAGGAAATGATGTCTCAGGCACTGAAAGCCACCTTCAGTGGCTTC 249

      135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
                :::  ||||:::|||      ||||||:::::::::
      250 GCAAAGGAGCAGCAGCGGCTGGGAATCCCCAAAGATCCCCAGCAGTGGAC 299

      151 rProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArqL 168
          :  ::::::|||:::  |||:::::|||:::  :::::::::  |
      300 AGAGACGCACGTGCGGGACTGGGTGATGTGGGCAGTGAATGAGTTCAGCC 349

      168 euProProMetGlyLysAlaPheGlnGlu......LeuAlaGlyLysGln 182
          ||        |||      ||||||:::      :::  |||
      350 TG......AAGGCAGTGGATTTCCAGAAGTTCTGCATGAACGGAGCTGCC 393

      183 LeuCysAlaMetSerGluGlnGlnPheArqGlnArqSerPro...LeuGl 198
          ||||||||||:::::|||  |||  :::      |||  :::
      394 CTGTGCGCCCTGGGCAAGGAGTGCTTCCTGGAGCTACGGCCTGACTTTGT 443

      198 yGlyAspValLeuHisAlaHisLeuAspIleTrpLys.............. 210
           ||||||:::|||      ||||||:::|||   :::
      444 GGGAGATATCCTTTGGGAACACCTGGAGATCTTGCAGAAAGAAGAGGCAA 493

      211 .....................SerAlaAlaTrpMetLysGluArq... 218
                                :::|||||:::   :::::|||
      494 AACCATACCCAGCAAATGGAGTGAATGCAGCGTATCCAGAATCCCGCTAT 543

      219 ThrSerProGlyAlaIleHisTyr.................CysAlaSe 229
          ||||||     |||   |||                    ||||::
      544 ACTTCAGACTACTTCATTAGTTATGGCATCGAGCACGCACAGTGCGTGCC 593

      229 rThrSerGlu.....................GluSerTrp..... 239
             ||||||                    |||||||:::
      594 TCCCTCCGAGTTCTCTGAGCCAGCTTCATCACAGACTCCTACCAGACCC 643

      235 ..................................................... 235

      644 TCCATCCCATCAGCTCGGAAGAGCTTCTGTCCCTCAAGTACGAGAACGAC 693

      236 ..................................ThrAspSer..... 238
                                           |||||||||
      694 TATCCCTCAGTCATCCTTCGTGACCCCGTCCAGACGGACTCCCTGCAGAC 743

      239 .....................GluVal.....AspSerSerCysS 245
                             ||||||     |||:::   |||
      744 AGACTACTTCACAATCAAGCAAGAAGTGGTAACGCCAGACAACATGTGCA 793

      245 erGlyGln..................................... 247
          |||:::
```

```
      794 TGGGACGTGCCAGTCGAGGTAAACTGGGTGCCCAGGACTCCTTTGAGAGC 843

      247 ..................................................... 247

      844 ATAGAGAGCTACGACAGCTGTGACCGCCTGACACAGTCCTGGAGCAGCCA 893

      247 ..................................................... 247

      894 GTCCTCCTTCCAGAGCCTGCAGCGCGTCCCCTCCTACGATAGCTTTGACT 943

      247 ..................................................... 247

      944 CAGAGGACTACCCCGCCGCCCTGCCCAACCACAAGCCCAAGGGCACCTTC 993

      247 ..................................................... 247

      994 AAGGACTATGTTCGACATCGGGCTGACATGAACAAGGACAAGCCTGTCAT 1043

      248 .....................................ProIleHisLeuI 262
                                              ||||||:::|||
     1044 TCCTGCCCCTGCCCTCGCCGGCTACACAGGCAGTGGACCCATCCAACTGT 1093

      263 leArgTrpLeuAsnLysGluLysGlyIlePheLysIleIleAspSerAr 282
           |||   |||   :::   :::          ||||||:::::|||
     1141 ATAAGCTGGACGGGT...GATGGCTGGGAGTTCAAGCTTTCCGATCCAGA 1187

      283 gGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnT 302
           :::||||||||||   ||||||   ||||||||||:::|||   |||||||
     1188 TGAGGTGGCCAGGCGGTGGGGCAAGAGGAAAAACAAGCCCAAGATGAACT 1237

      303 yrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIle 338
           |||:::|||||||||||||:::::|||   ||||||   |||:::||||||
     1238 ATGAGAAGCTCAGCCGTGGTCTGCGTTACTATTACGACAAGAACATCATC 1287

      339 ArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal 372
           ::|||         :::::|||   |||||||:::|||||
     1288 CACAAG...ACGGCGGCAAGCGCTACGTCTACGGCTTCGTC 1326

seq_name: qb ov:XLCETS2B

seq_documentation block:
LOCUS       XLCETS2B     2476 bp    mRNA            VRT       17 FEB 1997
DEFINITION  X.laevis mRNA for Xl-c ets-2b protein.
ACCESSION   X52645
VERSION     X52645.1  GI:64620
KEYWORDS    c-ets proto-oncogene; DNA binding protein.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 464 to 1747)
  AUTHORS   Stiegler,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (20 APR 1990) Stiegler P., Institut de Biologie
            Moleculaire et Cellulaire du CNRS, 15 rue rene Descartes, 67084
            Strasbourg Cedex, France
  REMARK    revised by [4]
REFERENCE   2  (bases 1 to 1285)
  AUTHORS   Wolff,C.M., Stiegler,P., Baltzinger,M., Meyer,D., Ghysdael,J.,
            Stehelin,D., Befort,N. and Remy,P.
  TITLE     Isolation of two different c-ets 2 proto-oncogenes in Xenopus
            laevis
  JOURNAL   Nucleic Acids Res. 18 (15), 4603-4604 (1990)
  MEDLINE   90356411
REFERENCE   3  (bases 1 to 1285)
  AUTHORS   Wolff,C.M., Stiegler,P., Baltzinger,M., Meyer,D., Ghysdael,J.,
            Stehelin,D., Befort,N. and Remy,P.
  TITLE     Cloning, sequencing, and expression of two Xenopus laevis c-ets-2
            proto-oncogenes
  JOURNAL   Cell Growth Differ. 2 (9), 447-456 (1991)
  MEDLINE   92088972
REFERENCE   4  (bases 1 to 2476)
  AUTHORS   Stiegler,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (06 NOV 1991) Stiegler P., Institut de Biologie
            Moleculaire et Cellulaire du CNRS, 15, rue Rene Descartes, 67084
            Strasbourg, CEDEX, France
COMMENT     Data kindly reviewed (11-SEP 1990) by Stiegler P.
            See also: X51826 for Xl c-ets-2a.
FEATURES             Location/Qualifiers
     source          1..2476
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /dev_stage="adult"
                     /tissue_type="ovary"
                     /cell_type="oocyte (stage I,II)"
                     /clone="c-1522, c-1511"
     CDS             103..1521
                     /codon_start=1
                     /product="Xl c-ets-2b protein"
                     /protein_id="CAA30860.1"
                     /db_xref="GI:64621"
                     /db_xref="SWISS-PROT:Q01712"
                     /translation="MTEEFGIRNMDQVAPVYNGHRGMLKRQLAFDNVQVPTSLYGLES
                     AYEEEGAVLPGLLEYPHDSSSCELPLLTPCSPAVMSQALKDTFNGFAKERCRLGIPGN
                     PWLWDENNVLQWLLWAAKEFSLENVNFQKFLMNGHELCSLGKERFLALADPDVGDTLW
                     EHLEEMMKEYQEKAQEPYTQHSNRPDSLNQWMNADSLNETADPLQCGAQVRNYFKNGMY
                     NDMCSVPTGQTLLNPKQEFQQYPSSCLKSRAVNYPPASQDFARSHMNVLLNSLNSGKL
                     RDYDSGDSGTESFESTESLLHSWTSQSSLVDMQRVPSYDSFEEDGNQTLCLNKQFMSF
                     KDYTQDRCEEFAELGKPVTLPASTLAGFTGSGFTQDLWQFLLELLTDKSCQSFISWTGDGW
                     EFKLADPGEVARRWGKRKNKPKMNYEKLSRGLRYYYDKNIIHKTSGKRYVYRFVCDLH
                     NLLGYTPDELHAMLGVQPDTDE"
     old_sequence    463..1747
                     /citation=[1]
     old_sequence    1745
                     /citation=[1]
                     /replace="a"
BASE COUNT      681 a    509 c    557 g    729 t
ORIGIN


alignment_scores:
            Quality:  260.00              Length:    458
              Ratio:  1.444                 Gaps:     17
 Percent Similarity:  89.301   Percent Identity:  22.707

alignment_block:
 QS 09-126-945B-2 x XLCETS2B  ..

 Align seq 1/1  to: XLCETS2B  from: 1  to: 2476

      65 MetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSe 81
         :::|||      :::|||      |||   :::|||          :::::
      90 CTGCTCAGGGCCAATGACAGAGTTTGGAATTCGAAACATGGACCAAGTGG 139

      81 rSerArgGluGluProProGluGluProGluGlnCysProValIleAspS 98
         :|||            |||:::::      |||:::||||||
     140 CTCCAGTGTATAACGGCCACAGAGGAATGCTCAAGCGCCAGTTACGCTTT 189

      98 erGlnAlaProAlaGlySerLeu..........................  105
          |||:::|||:::::::|||
     190 GACAATGTCCAAGTGCCGACCTCGCTCTACTGTGGCTTATTTTCTGCTTA 239

     106 .........Asp.LeuValProGlyGlyLeuThrLeuGluGluHisSer. 118
                  ||||||   ||||||          |||:::
     240 TGAAGAAGAACAGGCCGTGCCAACAGGCCTGGACTCTTACCCTCATGATT 289

     119 ......................LeuGluGlnValGlnSerMetValVal 127
                              |||      :::::   |||:::
     290 CCAGCAGCTGTGAGTTACCCCTGCTAACCCCGTGCAGCAAAGCTGTGATG 339
```

```
 128 GlyGluValLeuLysAsp.............IleGluThrAlaCysLy 159
     :::::::::||||||||               :::     |||::
 340 AGTCAAGCGCTGAAAGATACTTTAATGGTTTTGCGAAGGAACGATGCAG 389

 139 sLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsnValG 156
     :   |||:::|||   :::::::|||     |||:::   :::|||||
 390 A...CTTGGCATCCCCGGCAATCCATGGCTTTGGGATGACAATAATGTGT 436

 156 lnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGly 172
       :::|||||||||||||::       :::::   |||     :::::
 437 TACAGTGGCTTTTGTGGGCCGCCAAGGAATTCTCCTTGGAGAACGTGAAT 486

 173 LysAlaPheGlnGlu......LeuAlaGlyLysGlnLeuCysAlaMetSe 197
           ||||||::       :::   |||   ||||||||||::::::
 487 .....TTTCAGAAGTTTCTCATGAATGGACACGAGCTATGCAGCCTCGT 536

 197 rGluGluGlnPheArgGlnArgSerPro...LeuGlyGlyAspValLeuH 208
     ::::|||::|||          :::|||   :::   ||||||:::|||
 537 GAAGGAGAGGTTTTGGCTCTGGCACCTGACTTCGTTGGGGACATACTTT 586

 208 tsAlaHisLeuAsp..................................... 207
           |||||||::
 587 GGGAACATCTGGAAGAAATGATGAAAGACTATCAAGAAAACGCCCACGCAG 636

 208 .....IleTrpLysSerAl 212
                 |||   ::::|
 637 CCATATATTGACCATTCTAACCGGGACTCACTAAACCAGTGGATCAATGC 686

 212 aAlaTrpMetLysGluArgThrSerPro..................... 221
     |      ::::::      :::::|||
 687 AGATTCCTTAAATTTCACTGCTGATCCGTTGCAGTGTGGGGCACAAGTAC 736

 221 ..................................................... 221

 737 ACAATTACCCTAAAAATGGAATGTATAATCATATGTCCTCACTGCCCACT 786

 221 ..................................................... 221

 787 GGTCAGACTCTCCTCAACCCCAAACAAGAGTTCCAACAATACCCCAGCTC 836

 222 .............GlyAlaIleHisTyrCysAlaSerThrSerGlu.... 232
                 |||::::::|||      ::::::::::::
 837 TTGCCTAAAATCCAGAGCAGTAAACTACCCGCCCGCCAGCCAGGACTTG 886

 232 ..................................................... 232

 887 CAAGGAGCCATATGAACGTGTTGCTCAACTCTCTAAATTCTGGCAAATTG 936

 232 ..................................................... 232

 937 AGAGACTATGACTCAGGAGACAGCGGCACAGAAAGCTTCGAGAGCACCGA 986

 233 .........GluSerTrpThrAspSerGlu.................... 239
             :::|||||||||||:::::::
 987 ATCACTACTGCATTCTTGGACGAGCCAATCTTCTCTTGTCGATATGTAAC 1036

 239 ..................................................... 239

1037 GAGTTCCTTCATATGACAGCTTTGAGGAAGATGGCAATCAAACGTTGTGT 1086

 240                                   ValAspSerSerCys   251
                                       :::::::::   |||
1087 CTAAATAAGCAACCAATGTCCTTCAAAGATTATATCCAGGACAGGTGTGA 1136

 244 ..................................................... 244

1137 GCCAGCAGAACTAGGCAAACCTGTTATTCCCGCATCTATACTTGCTGGCT 1186

 245 .........SerGlyGlnPr-IleHisLeuTrpGlnPheLeuLysGluLeu 258
             ||||||   ||||||:::|||||||||||||||||   |||
1187 TCACAGGCAGTGGA...CCGATCCAATTGTGGCAGTTCTTGTTGGAA... 1224

 259 LeuLeuLysProHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGl 275
```

```
     ||||||          |||          ||||||   |||   :::   ::
1225 CTTTTAACGGATAAGTCATGTCAGTCCTTATCAGCTGGACTGGG...GA 1271

 275 uLysGlyIlePheLysIleGluAspSerAlaGlnValAlaArgLeuTrpG 292
     :          |||||:::   |||       :::||||||||   ||||
1272 TGGATGGGAGTTCAAGCTGGCTGACCCAGATGAGGTGGCCCGGCGGTGGG 1321

 292 lyIleArgLysAsnArgProAlaMetAsnTyrAspLysLeuSerArgSer 308
     ||   ||||||||||:::|||   ||||||||||:::||||||||||:::
1322 GTAAAAGGAAAAACAAACCTAAAATGAACTACGACAAGCTAAGTCGACGA 1371

 309 IleArgGlnTyrTyrLysLysGlyIleIleArgLysProAspIleSerGl 325
     :::|||   ||||||   |||:::||||||:::|||   :::   :::::
1372 CTAAGGTATTATATCACAAGAACATCATCCACAAGACATCA...GGGAA 1418

 325 nArgLeuValTyrGlnPheVal 332
     :|||   ||||||:::||||||
1419 GAGATACGTGTACCGATTTGTT 1440

seq_name: gb_pr5:AF256222

seq_documentation_block:
LOCUS       AF256222     2658 bp    mRNA            PRI       02 MAY 2000
DEFINITION  Homo sapiens ets family transcription factor ELF2A (ELF2) mRNA,
            complete cds, alternatively spliced.
ACCESSION   AF256222
VERSION     AF256222.1  GI:7677542
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2658)
  AUTHORS   Wilkinson,D.A., Neale,G.A., Mao,S., Naeve,C.W. and Goorha,R.M.
  TITLE     Elf-2, a rheumatoid arthritis synovial cell transcription factor: discovery
            and potential role in T cell leukemia
  JOURNAL   Leukemia 11 (1), 86-96 (1997)
  MEDLINE   97154704
REFERENCE   2  (bases 1 to 2658)
  AUTHORS   Wilkinson,D.A., Neale,G.A.M., Mao,S., Fernandes,E.R.,
            Davenport,J.W., Naeve,C.W. and Goorha,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-2000) Virology and Molecular Biology, St. Jude
            Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38105,
            USA
FEATURES             Location/Qualifiers
     source          1. .2658
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /cell_line="K562"
     gene            1. .2658
                     /gene="ELF2"
     CDS             183. .1928
                     /gene="ELF2"
                     /note "alternatively spliced; includes N terminal
                     transactivation domain"
                     /codon_start=1
                     /product "ets family transcription factor ELF2A"
                     /protein_id "AAF67195.1"
                     /db_xref "GI:7677543"
                     /translation="MESAVVSGCTLVFSSNCVENGFPSPPVSFVFDAVIVFSVCAR
                     LFQGYAAQVIVYDDETYMMQDVAEEQEVETIENVPETVEASVHSSNAHCTLDKTIEAAEAL
                     LHMESPTCLPDSPSPFFTHAAMPPDVLTPTVVFVSIFFSPFPMDISPTPTSPDSHFFMK
                     KFFVCPPPFTQQSPISNGSPSLCTKKKPPPCKGNTTYYIWFFLIDLLQKRNICDRYTKW
                     TQREKGIFKLVDSKAVSKLWGKHSNKPDMNYETMSRALRYYYDRGIIAKVEGQRLVYQ
                     FKDMPKNIVVIHDDKSFTCNPDLAGTIDFKSIFRVSLSACSLCKAASSVRSGKNSSPI
                     NCSPAFFGVAKVVNITSPGHDASSPSPTTTASVSAIAAERTVRVAMQVVVMISTCQK
                     ISTVAVQSVNAGAPLITSTSPTTATSPKVVIQTIPTVMPASTENGDKTTMQPAKTTTI
                     PATQLAQQQLQTKSNETGSGSINIVGTPIAVRALTPVSIAHETPVMETSMPTQQASGL
                     TPPPVTSAVIEGFEVESEAVAKEQEHDVKTIQEVEEKPAIGNKTVTHVVVVSAPSAIA
                     LPVTMKTEGLVICEK"
     misc_feature    888. .1016
                     /gene "ELF2"
```

```
                     /note "ets domain; DNA binding site"
     misc_feature    1452..1718
                     /gene "ELF2"
                     /note "LMO2 binding site; DNA binding site"
BASE COUNT       917 a    507 c    551 g    678 t
ORIGIN


alignment scores:
            Quality:  260.00              Length:     345
              Ratio:  1.585                 Gaps:      11
 Percent Similarity:  47.536   Percent Identity: 24.058

alignment block:
 US-09-126-945B-2 x AF256222   ..

 Align seq 1/1  to: AF256222  from: 1  to: 2653

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa  66
         |||:::  ::::::  |||:::          ::::::|||:::::
     168 CCTCAGGGAATAACAATGACATCAGCAGTGGTTGACAGTGGAGGTTACTAT 217

      66 l.................GlyLeuGluArgArgAspTrpSer.......  84
         :                 |||:::|||::::::::::   |||
     218 TTTGGAGCTTTCCAGCAATGGAGTAGAAAATCAAGAGGAAAGTGAAAAGG 267

      85 ..............................ProSerProProAlaThrPro  51
                                      |||  |||  |||
     268 TTTCTGAATATCCAGCAGTGATTGTGGAGCCAGTTCCAAGTGCCAGATTA 417

      52 GluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyr..  67
         |||||||||   :::|||     |||  |||  |||     |||
     418 GAGCAGGGCTATGCAGCCCAGGTTCTGGTTTATGATGATGACACTTATAT 367

      68 .ProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgG  84
          :::|||   :::                                   |
     368 CATGCAAGATGTGGCA................................G  384

      84 luGluProProGluGluProGluGlnCysProValIleAspSer......  98
         |||||         |||  |||:::      ::::::::::::
     385 AAGAACAAGAAGTTGAGACCGAGAATGTGGAAACAGTGGAAGCATCAGTT 434

      99 .....................................GlnAlaProAlaGl 103
                                            :::|||      ::
     435 CACAGCAGTAATGCACACTGTACAGATAAGACAATTGAAGCTGCTGAAGC 484

     103 ySerLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuG 120
         :  |||  :::                        :::::::::|||   |
     485 CCTGCTTCATATGGAATCTCCTACCTGCTTGAGGGATTCAAGAAGTCCTG 534

     120 luGlnValGlnSerMetValValGlyGluValLeuLysAspIleGluThr 136
         ||  :::::::::  :::      :::|||:::         ||||||
     535 AATTCATCCATGCTGCTATCAGGCCAGATGTCATT......ACAGAAACT 578

     137 AlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSe 153
         :::  ::::::  :::      ::::::|||||||   ||||||
     579 GTAGTGGAGGTGTCAACTGAAGAGTCTGAACCCATGGATACCTCTCCT... 626

     153 rAsnValGluLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProP 170
                                            :::|||
     627 ..........................................ATTCCAA 644

     170 roMetGlyLysAlaPheGlnGlnLeuAlaGlyLysGluLeuCysAlaMet 186
         :::  :::  :::      |||:::          :::
     644 CATCACCAGATAGCCATGAACCAATGAAAAAGAAAAAA........GTT 674

     187 SerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValLeuHi 203
         :::::::::          |||:::|||||:::::::::
     675 GGCCGTAAACCAAAGACCCAGCAATCACCAATTTCCAATGGG........ 716

     203 sAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgIhrS 220
                            :::          :::|||::::::
     717 ..................TCTCCTGAGTTAGGTATAAAGAAGAAACCAA 747

     220 orProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpThr 236
             |||
     748 GAGAAGCA.............................................. 755

     237 AspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGl 254
                                    :::|||:::      ::::::::::
     756 .......................AAAGGAAACACAACCTATTTGTGGGA 781

     255 nPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleA 270
         :||||||   :::|||||||          ::::::      |||:::|||:
     782 GTTTCTTTTAGATCTACTTCAAGATAAAAATACTTGTCCCAGGTATATTA 831

     270 rgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGln 286
         ..|||   ......|||||||||||||||||...   ||||||
     832 AATGGACTCAGAGAGAAAAAGGCATATTCAAGCTGGTGGATTCAAAGGCT 881

     287 ValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAs 303
         |||::::::::|||||||||   :::||||||:::|||   ||||||||||:
     882 GTCTCTAAGCTTTGGGGAAAGCATAAGAACAAACCAGACATGAACTATGA 931

     303 pLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgL 320
         :   :::::::|||:::::::|||   ||||||:::::|||||:::   |
     932 AACCATGGGACGAGCTTTGAGATACTACTACCAAAGGGGAATTCTTGCAA 981

     320 ysProAspIleSerGlnArgLeuValTyrGlnPhe 331
         ||   :::   :::||||||||||||||||||||||
     982 AGGTTGAA...GGACAGAGGCTTGTATATCAGTTC 1013

seq_name: qb_pri0:HS043188

seq_documentation_block:
LOCUS        HS043188      3222 bp    mRNA            PRI       05-DEC-1996
DEFINITION   Human Ets transcription factor (NERF-2) mRNA, complete cds.
ACCESSION    043188
VERSION      043188.1  GI:1420888
KEYWORDS     .
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 3222)
  AUTHORS    Oettgen,P., Akbarali,Y., Boltax,J., Best,J., Kunsch,C. and
             Libermann,T.A.
  TITLE      Characterization of NERF, a novel transcription factor related to
             the Ets factor ELF-1
  JOURNAL    Mol. Cell. Biol. 16 (9), 5091-5106 (1996)
  MEDLINE    96447578
REFERENCE    2  (bases 1 to 3222)
  AUTHORS    Libermann,T.A., Oettgen,P., Kunsch,C., Akbarali,Y. and Boltax,J.
  TITLE      Direct Submission
  JOURNAL    Submitted (13-DEC-1995) Towia A. Libermann, Medicine, Beth Israel
             Hospital, 330 Brookline Avenue, Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..3222
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="spleen; fetal liver; fetal brain"
     gene            207..1952
                     /gene="NERF-2"
     CDS             207..1952
                     /gene="NERF-2"
                     /codon_start=1
                     /product="Ets transcription factor"
                     /protein_id="AAB37759.1"
                     /db_xref="GI:1420889"
                     /translation="MTSAVVIEGGTTLEELSNRGVENGEESEEKVSLYPAVIVEEVPSAR
                     LEQGYAAQVLVYDDETYMMQDVAEEQEVETENVETVEASVHSSNAHCTDKIIEAAEAL
                     LHMESPTCLRDSRSPEFIHAAMRPDVTTETVVEVSTEESEPMDTSPTIPTSPDSHEPMK
                     EKEAVGREKPKIQQQSPTSNGSPEEGTKEKPPEFGKGNLIYLWEELLPLLQQQKNICEGYLKW
                     IQKEKGTFEELVDSEAVSKLWGAHKNKPDMNYEIMSLPALRYYYQRGILAKVEAQQKLVYQ
                     EKDMPKNIVVIDDDKSETGNEDLACTIDEFSLEPVSLSAESLIKAASSVRSGKNSSPI
                     NCSRAEKGVARVVNITSPGHDASSRSPTTTASVSATAAPRTVRVAMQVPVVMTSLGQK
```

```
                    ISIVAVQSVNAGAPLITSISPTTAISPEVVIQTIPTVMPASIENCLKTIMQFAKTIII
                    PAIQLAQCQLQTKSNLIGSGSINIVGTPLAVRALTPVSTARTTPVMRLSMPTSQAVSK
                    TPPRVISAVIKGPLVKSEAVAKEQERHQVKTIQLVLKPAGNSTVTHVVVVSAPSAIA
                    LPVTMKTEGLVTCEK"
BASE COUNT     1055 a    592 c    658 g    917 t
ORIGIN


alignment_scores:
            Quality:  260.00              Length:     345
              Ratio:  1.585                 Gaps:      11
 Percent Similarity:  47.536   Percent Identity: 24.058

alignment_block:
 US 09-126-945B-2 x HSU43188  ..

 Align seq 1/1  to: HSU43188  from: 1  to: 4222

      20 ProAspThrValSerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaVa  66
         |||:::  :::::  |||:::           :::::|||:::::
     192 CCTCAGGGAATAACAATGACATCAGCAGTGGTTGACAGTGGAGGTACTAT 241

      66 l................GlyLeuGluArgArgAspTrpSer.......  44
         :                |||:::|||:::::::::  |||
     242 TTTGGAGCTTTCCAGCAATGGAGTAGAAAATCAAGAGGAAAGTGAAAAGC 291

      45 ..........................ProSerProProAlaThrPro  51
                                   |||  |||  |||
     292 TTTCTGAATATCCAGCAGTGATTGTGGAGCCAGTTCCAAGTGCCAGATTA 341

      52 GluGlnGlyLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyr..  97
         |||||||||  :::|||      |||  |||  |||      |||
     342 GAGCAGGGCTATGCAGCCCAGGTTCTGGTTTATGATGATGAGACTTATAT 391

      98 .ProGluAspSerSerTrpAlaAlaLysAlaProGlyAlaSerSerArgS  84
          :::|||  :::                                    |
     392 GATGCAAGATGTGGCA................................G 408

      84 luGluProProGluGluProGluGlnCysProValIleAspSer.....  98
         |||||        |||  |||:::  ::::::::::::
     409 AAGAACAAGAAGTTGAGACCGAGAATGTGGAAACAGTGGAAGCATCAGTT 458

      99 .....................................GlnAlaProAlaGl 103
                                             ::::||      ::
     459 CACAGCAGTAATGCACACTGTACAGATAAGACAATTGAAGCTGCTGAAGC 508

     104 ySerLeuAspLeuValProGlyGlyLeuThrLeuGluGlaHisSerLeuS 120
         :  |||  :::                 ::::::::::||  |
     509 CCTGCTTCATATGGAATCTCCTACCTGCTTCAGGGATTCAAGAAGTCCTG 558

     120 lnGlnValGlnSerMetValValGlyGluValLeuLysAspIleGlaThr 166
         ||  :::::::::  :::      :::|||:::          ||||||
     559 AATTCATCCATGCTGCTATGAGGCCAGATGTCATT......ACAGAAACT 602

     167 AlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSe 213
         :::  ::::::  :::      ::::::||||||||  ||||||
     603 GTAGTGGAGGTGTCAACTGAAGAGTCTGAACCCATGGATACCTCTCCT.. 650

     153 rAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnIyrArqLeuProP 170
                                                 ::::||
     651 ..........................................ATTCCAA 657

     170 roMetClyLysAlaPheGlnGluLeuAlaClyLysGluLeuCysAlaMet 196
                 :::  :::  :::      ||:::          :::
     658 CATCACCAGATAGCCATGAACCAATGAAAAAGAAAAAA.........GTT 698

     187 SerGluGluGlnPheArqGlnArqSerProLeuGlyGlyAspValLeuHi 203
         :::::::::          ||:::|||||||:::::::::
     699 GGCCGTAAACCAAAGACCCAGCAATCACCAATTTCCAATGGG........ 740

     203 sAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGluArgThrS 220
                       :::          :::|||:::::::
     741 ..................ICTCCTGAGTTAGGTATAAAGAAGAAACCAA 771

     220 erProGlyAlaIleHisTyrCysAlaSerThrSerGlnGlnSerTrpTh 236
             |||
     772 GAGAAGGA..................................... 779

     237 AspSerGluValAspSerSerCysSerGlyGlnProIleHisLeuTrpGl 253
                                  :::|||:::      :::||||||:
     780 .......................AAAGGAAACACAACCTATTTGTGGGA 805

     253 nPheLeuLysGluLeuLeuLysProHisSerIyrGlyArgPheIleA 270
         ::||||   :::||||||         ::::::      |||:::|||
     806 GTTTCTTTAGATCTACTTCAAGATAAAAATACTTGTCCCAGGTATATTA 855

     270 rqTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGln 286
         ::|||   :::::|||||||||||||:|||||:::   ||||||
     856 AATGGACTCAGAGAGAAAAAGGCATATTCAAGCTGGTGGAATTCAAAGGCT 905

     287 ValAlaArgLeu:rpGlyIleArgLysAsnArgProAlaMetArgIyrAs 333
         |||:::::|||||||||   :::|||||:::::||   ||||||||||:
     906 GTCTCTAAGCTTGGGGAAAGCATAAGAACAAACCAGACATGAACTATGA 955

     303 pLysLeuSerArgSerIleArgGlnIyrTyrLysLysGlyIleIleArgL 320
         :  :::::::|||:::::|||   |||||:::::::|||||:::   |
     956 AACCATGGGACGAGCTTTGAGATACTACTACCAAAGGGGAATTCTTGCAA 1005

     320 ysProAspIleSerGlnArqLeuValIyrGlnPhe 354
         ||   :::   :::|||||||||:|||||||||
    1006 AGGTTGAA...GGACAGAGGCTTGTATATCAGTTC 1037

seq_name: qb_pri10:HSU11732

seq_documentation_block:
LOCUS       HSU11732     1580 bp    mRNA            PRI       17 JUL 1994
DEFINITION  Human ets like gene (tel) mRNA, complete cds.
ACCESSION   U11732
VERSION     U11732.1  GI:511282
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1580)
  AUTHORS   Golub,T.R., Barker,G.F., Lovett,M. and Gilliland,D.G.
  TITLE     Fusion of PDGF receptor beta to a novel ets-like gene, tel, in
            chronic myelomonocytic leukemia with t(5;12) chromosomal
            translocation
  JOURNAL   Cell 77 (2), 307-316 (1994)
  MEDLINE   94221647
REFERENCE   2  (bases 1 to 1580)
  AUTHORS   Gilliland,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1994) D. Gary Gilliland, Hematology/Oncology,
            Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115,
            USA
FEATURES             Location/Qualifiers
     source          1..1580
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="tel-452"
                     /clone_lib="K562 lambda gt11"
                     /map="12p13"
                     /chromosome="12"
                     /cell_line="K562"
                     /tissue_type="leukemia"
     gene            25..1488
                     /gene="tel"
     CDS             25..1488
                     /gene="tel"
                     /note="t(5;12) translocation breakpoint occurs after
                     nucleotide 487"
                     /codon_start=1
                     /protein_id="AAA19786.1"
                     /db_xref="GI:511283"
```

```
                     /Translation "MSETPAQCSIKQERISYTPPESPVPSYASSTPLHVPVPRALRME
                     EDSTRLPAHLRLQPIYWSKDDVAQWLKWAENEESLRPTDSNTFEMNGKALLLTKEDF
                     RYRSPHSGDVLYELLQHILKQRKPRTLESPEFHPGNSTHTQPEVTLHQNHEEDNCVQR
                     TPRPSVDNVHHNPPTTELLHRSRSPITTNHRDSPDPPQRPLPSPLDNMTRRLSPAERA
                     QGPRPHQENNHQESYPLSVSPMENNHCPASSESHPKPSSPRQESTRVTQLMPSPTMHP
                     LTLNPRHSVDFKQSRLSEDGLHREGKPTNLSHREDLAYMNHTMVSVSPPEEHAMPTGR
                     TADCRLLWDYVYQLLSDSKYENFTRWEEKESKTFRTVDPNGLARLWQNHKNRTNMTYE
                     KMSKALRHYYKQNTTRKEPGQKLLFRFMKTPDETMSGRTDRLEHLESQELDRQTYQED
                     EC"
BASE COUNT       419 a    471 c    386 g    404 t
ORIGIN


alignment scores:
            Quality:  259.50              Length:    435
              Ratio:  1.653                 Gaps:     11
 Percent Similarity:  36.092   Percent Identity: 21.149

alignment block:
 US-09-126-945B-2 x HSU11732   ..

 Align seq 1/1  to: HSU11732  from: 1  to: 1580

      86 ProProGluGluProGluGlnCysProValIleAspSerGlnAlaProAl 102
         ||||||||||:::|||           :::   |||   |||   ::
      79 CCTCCAGAGAGCCCA.............GTGCCGAGTTACGCTTCCTC 113

     102 aGlySerLeuAspLeu...ValProGlyGlyLeuThrLeuGluGlnHisS 118
         :   |||   :::   ||||||   :::|||   :::|||||||   |
     114 GACGCCACTTCATGTTCCAGTGCCTCGAGCGCTCAGGATGGAGGAAGACT 163

     118 erLeuGluGlnValGlnSerMetValValGlyGluValLeuLysAspIle 134
         |||||||
     164 CGATCCGCCTGCCT..................................... 177

     135 GluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSe 151
                            ::::::      :::|||:::   |||||
     178 ..................GCGCACCTGCGCTTGCAGCCAATTTACTGGAG 209

     151 rProSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGluTyrArgL 168
         |   ::::::|||   :::||||||   |||:::|||:::::::::   |
     210 CAGGGATGACGTAGCCCAGTGGCTCAAGTGGGCTGAAAATGAGTTTTCTT 259

     168 euProProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCys 184
         ||   |||:::  :::          |||:::   ||||||   |||
     260 TAAGCCCAATTGACAGCAACACGTTTGAAATGAATGGCAAAGCTCTCCTG 309

     185 AlaMetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspVa 201
         :::::::::|||:::|||||||   |||||||||   :::|||||||||
     310 CTGCTGACCAAAGAGGACTTTCGCTATCGATCTCCTCATTCAGGTGATGT 359

     201 IleuHis.........AlaHisLeuAspIleTrpLysSerAlaAlaTrpM 215
         ||||:::             |||:::                  :
     360 GCTCTATGAACTCCTTCAGCATATT.....................C 389

     215 etLysGluArg..................................... 218
         ::|||:::|||
     390 TGAAGCAGAGGAAACCTCGGATTCTTTTTCACCATTCTTCCACCCTGGA 439

     218 ..................................................... 218

     440 AACTCTATACACACACAGCCGGAGGTCATACTGCATCAGAACCATGAACA 489

     218 ..................................................... 218

     490 AGATAACTGTGTCCAGAGGACCCCCAGGCCATCCGTGGATAATGTGCACC 539

     218 ..................................................... 218

     540 ATAACCCTCCCACCATTGAACTGTTGCACCGCTCCAGGTCACCTATCACG 589

     218 ..................................................... 218

     590 ACAAATCACGGGCCTTCTCCTGACCCCGAGCAGCGGCCCCTCCGGTCCCC 639

     218 ..................................................... 218

     640 CCTGGACAACATGATCCGCCGCTCTCCCCGGCTGAGAGAGCTCAGGAC 685

     219 .....................................................Thr 219
                                                              :::
     686 CCAGGCCGCACCAGGAGAACAACCACCAGGAGTCCTACCCTCTGTCAGTG 735

     220 SerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrpTh 236
         ||||||         ::::::|||   ::::::||||||::
     736 TCTCCCATGGAGAATAATCACTGCCCAGCGTCCTCGGAGTCCCACCCGAA 785

     236 rAspSer.............................................. 238
             |||
     786 GCCATCCAGCCCCCGGCAGGAGAGCACACGCGTGATCCAGCTGATGCCCA 835

     239 .....................................GluValAspSer 242
                                             :::||||||
     836 GCCCCATCATGCACCCTCTGATCCTGAACCCCCGGCACTCCGTGGATTTC 885

     243 SerCysSer.............................GlyGlnProIl 249
         :::   |||                            |||:::|||||
     886 AAACAGTCCAGGCTCTCCGAGGACGGGCTGCATAGGGAAGGGAAGCCCAT 935

     249 eHis.................................................. 250
         ||::
     936 CAACCTCTCTCATCGGGAAGACCTGGCTTACATGAACCACATCATGGTCT 985

     250 ..................................................... 250

     986 CTGTCTCCCCGCCTGAAGAGCACGCCATGCCCATTGGCAGAATAGCAGAC 1035

     251 .........LeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSe 264
                 ||||||||||||||   :::   |||||||::
    1036 TGTAGACTGCTTTGGGATTACGTCTATCAG...TTGCTTCTGACAGCCG 1082

     264 rTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLys1 281
         |||   :::|||||||||||||   :::||||||:::   ||||||:::|
    1083 GTACGAAAACTTCATCCGATGGGAGGACAAAGAATCCAAAATATTCCGGA 1132

     281 leGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArg 297
         ||   |||          :::|||||||||||||||   :::||||||||
    1133 TAGTGGATCCCAACGGACTGGCTCGACTGTGGGGAAACCATAAGAACAGA 1182

     298 ProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLy 314
         ||:::|||:::|||:::||||||:::::||:::||||||||
    1183 ACAAACATGACCTATGAGAAAATGTCCAGAGCCCTGCGCCACTACTACAA 1232

     314 sLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrGlnP 331
         |   :::||||||||||||   :::   :::|||||||:::::::::|
    1233 ACTAAACATTATCAGGAAG...GAGCCAGGACAAAGGCTTTTGTTCAGGT 1279

     331 heVal 332
         |||::
    1280 TTATG 1284

seq_name: gb_pri5:AF212848

seq documentation block:
LOCUS       AF212848     2047 bp    mRNA            PRI       02 APR 2000
DEFINITION  Homo sapiens Ets domain transcription factor (ESEJ) mRNA, complete
            cds.
ACCESSION   AF212848
VERSION     AF212848.1  GI:7385048
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2047)
  AUTHORS   Brooks Wilson,A.R., Buckler,A., Carden,I., Carey,A.H., Calvin,M.,
```

```
            Miller,A. and North,M.
  JOURNAL   Patent: PCT WO 99/37809 29-JUL-1999;
REFERENCE   2  (bases 1 to 2047)
  AUTHORS   Fuqores,A., Le,J., Sorokina,I., Snijders,A.J., Dayao,M.,
            Reddy,P.S., Mushegian,A., Carlee,L., Watanaskul,I., Chu,S.,
            Emtage,S. and McCormick,M.K.
  TITLE     A Novel Epithelial Specific ETS Transcription Factor, ESEJ, is a
            Negative Modulator of MAP Kinase Signaling Pathways, and its
            Expression is Lost in Carcinomas
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2047)
  AUTHORS   Fuqores,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1999) Dept. Functional Genomics, Axys
            Pharmaceuticals, Inc., 11099 N. Torrey Pines Rd., Suite 160, La
            Jolla, CA 92037, USA
FEATURES             Location/Qualifiers
     source          1. .2047
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "11"
                     /map "11p13"
     gene            1. .2047
                     /gene "ESEJ"
     CDS             132. .1034
                     /gene "ESEJ"
                     /note "EseJ"
                     /codon_start 1
                     /product "ETS domain transcription factor"
                     /protein_id "AAF61670.1"
                     /db_xref "GI:7385049"
                     /translation "MILPGGGVMNELNPGNNLLHQPPAWIDSYSTCNVSSGFPGGQWHE
                     IHPQYWTKYQVWWFWLQHLLDFNQLIDANCIPPQFFDINGPLCSMSLQFFTRAASIACQ
                     LLYLGNLQLHLKWNGQCSSSLFQCTHNVIVKIESEPCTMNIWEGNTEYIETNYGSIVDL
                     LDSFIPCPAVISMILESHLPVAESPEMRNLQDFASCLTSHNERGLTHLWTLTGCL
                     NPDKNPGLIRWFDPSPGVFSFLKSFAVAQIWGKFFNNSMIYLPISPAMPYYYSRFIL
                     ERVDGRRLVYKFGKNARGWRENEN"
BASE COUNT       603 a    455 c    473 g    516 t
ORIGIN


alignment_scores:
            Quality:  258.00              Length:    361
              Ratio:  1.410                 Gaps:     18
 Percent Similarity:  50.693   Percent Identity: 25.762

alignment_block:
 US-09-126-945B-2 x AF212848   ..

 Align seq 1/1  to: AF212848  from: 1  to: 2047

      43 TrpSerProSerProProAlaThrProGluGlnGly.....LeuSe 56
         |||  |||  |||  ::::::  |||  :::      |||
      26 TGGTGCCCCATCCCTATAGGAGCTGGTGAGATTGCAGCCTGCTGCCTCCC 75

      56 rAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGlu..... 69
         ::::::     |||||||||:   :::   :::|||||
      76 CTCCATCAGCCACAGCTATTGG...ATTTCCCACCCAGAATCTTTAGGTA 122

      70 ..........AspSerSer......TrpAlaAlaLysAlaProGlyAla 80
                  |||||||::      ||||::   :::        :::
     123 AATGAGATCATGATTCTGGAAGGAGGTGGTGTAATGAA.....TCT 163

      81 SerSerArgGluGluProProGluGluProGluGlnCysProValIleAs 97
         :::   |||:::::::||||||
     164 CAACCCCGGCAACAACCTCGT..........................  184

      97 pSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrL 114
            :::|||||||:::||||||||||
     185 .....TCACCACCCGCCAGCCTGGAC.....................  205

     114 enGluGlu.HisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
         :::::: ::|||      :::|||:::|||          |||
```

```
     206 ..AGACAGCTACTCCACGTGCAATGTTCCAGTGGGTTTTTTGGAGGCCA 253

     130 IleLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspP 147
            :::|||:::                                   |
     254 GTGGCATGAAATTCAT.................................C 270

     147 roMetAspTrpSerProSerAsnValGlnLysTrpLeuLeuTrpThrGlu 163
         ||:::   |||:::       :::|||   :::|||||
     271 CTCAGTACTGGACCAAGTACCAGGTGTGGGAGTGGCTCCAGCACCTCCTG 320

     164 HisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu......Le 178
                 ::::||                 |||||||||       ::
     321 GACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTTCGACAT 370

     178 nAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArgS 195
         :  |||::::::||||||:||||||   ::::::|||   :::   :
     371 CAACGCCGAGCACCTCTGCAGCATGAGTGAGGAGTTCACCCGGGCGG 420

     195 erProLeuGlyGlyAspValIleHisAlaHisLeu.............. 206
         ::      :::|||::::::|||:::::::::|'|
     421 CAGGGACGCCGGGGCAGCTCCTCTACAGCAACCTGCAGCATCTGAAGTGG 470

     207 ..................AspIleTrpLysSer.........AlaAlaTr 214
                           |||:::::::::::|||         :::
     471 AACGGCCAGTGCAGTAGTGACCTGTTCCAGTCCACACACAATGTCATTGT 520

     214 pMetLysGluArgThrSerProGlyAlaIle..................H 225
            |||::::|||::|||:::   :::                    :
     521 CAAGACTGAACAAACTGAGCCTTCCATCATGAACACCTGGAAAGACGAGA 570

     225 isTyrCysAlaSerThrSerGluGlnSerTrpThrAspSerGluValAsp 241
         ::|||      :::|||:::      |||   ::|||      :::|||
     571 ACTATTTATATCACACACAACTATGGTAGCACAGTAGATTTG...TTGGAC 617

     242 SerSer.....CysSerGlyGln..........................  247
         |||:::      |||   :::|||
     618 AGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAGTCACCT 667

     248 .Profile.....................................  249
          ||||::
     668 TCCTGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCCCTGCCA 717

     250 .....................................HisLeuTrpGlnPhe 254
                                            |||||||||:|||
     718 AGTGCCACACCAAAAAGCACAAGCCCACACCCACTCACTTATCCCAATTC 767

     255 LeuLysGluLeuLeuLeulysProHisSerTyrGlyArgPheIleArgLr 271
         ::::::::::::|||||:::|||   :::          :::|||:::||
     768 ATCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATAAAATG 817

     271 pLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnValA 288
         |   :::::::::::|||:::|||:::::::      |||      ||||
     818 GGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGCAGTGG 867

     288 laArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLys 304
         |:::||||||||||   :::|||||||::   :::|||:::|||:::|||
     868 CTCAGCTATGGGTAAAAAGAAGAACAACAGCAGCATGACCTATGAAAAG 917

     305 LeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPr 321
         ||||||||||:::::|||   ||||||||||::   |||:::::::::
     918 CTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATACTGGAGCGTGT 967

     321 oAspIleSerGlnArgLeuValTyrGlnPhe 341
          |||   :::::|||||||||||||:::|||
     968 GGAT...GGACGAAGACTGGTATATAAATTT 995

seq_name: qb_pr5:AF203977

seq_documentation_block:
LOCUS       AF203977     3160 bp    mRNA            PRI       01-DEC-2000
DEFINITION  Homo sapiens ETS family transcription factor EHF (EHF) mRNA,
            complete cds.
ACCESSION   AF203977
```

```
VERSION     AF203977.1  GI:11493772
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4160)
  AUTHORS   Mitchell,S.C., Levin,E., Hubert,R., Yeramian,C., Saffran,D.C. and
            Afar,D.E.H.
  TITLE     Up Regulation of the Epithelial-Specific Transcription Factor Genes
            hEHF and ESX in Advanced Prostate Cancer
  JOURNAL   Prostate (1999) In press
REFERENCE   2  (bases 1 to 4160)
  AUTHORS   Mitchell,S.C., Levin,E., Hubert,R., Yeramian,C., Saffran,D.C. and
            Afar,D.E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10 NOV 1999) DroGenesys Inc., 1701 Colorado Ave., Santa
            Monica, CA 90404, USA
FEATURES             Location/Qualifiers
     source          1..4160
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p13"
     gene            1..4160
                     /gene="EHF"
     CDS             113..1015
                     /gene="EHF"
                     /codon_start=1
                     /product="ETS family transcription factor EHF"
                     /protein_id="AAG35644.1"
                     /db_xref="GI:11493773"
                     /translation="MTLEGGGVMNLNPGNNLLHQIPAWTDSYSTCNVSSGFPGGQWHE
                     THPQYWTKYQVWEWLQHLLDTNQLDANCIPFQEFDINGEHLCSMSLQEFTRAAGTAQQ
                     LLYSNLAHLKWNGQLCSSDEFQSTHNVIVRTPSTMNTWKDENYLYFFNYGLSIVDL
                     LDSKTFCKAQTSMFTSHLPVAFSPDMFKEQDPPAFCHLEKHNPPGTHLWRFFRDLLL
                     NFDENPSGLEWEDWSEGVEFFLESLAVAQLWGKEENNSSMSYEKLSFAMFYYYKFEIL
                     ERVDGKRLVYKFGKNARGWRENEN"
BASE COUNT       942 a     634 c     683 g     838 t      63 others
ORIGIN


alignment scores:
            Quality:  258.00              Length:     361
              Ratio:  1.410                 Gaps:      18
 Percent Similarity:  56.693   Percent Identity: 25.762

alignment block:
 DS:09 126 945B 2 x AF203977  ..

 Align seq 1/1  to: AF203977  from: 1  to: 4160

      43 TrpSerProSerProProAlaThrProGluGlnGly.....LeuSe 56
         |||  |||  |||  ::::::  |||  :::        |||
       7 TGGTGCCCCATCCCTATAGGAGCTGGTGAGATTGCAGCCTGCTGCCTCCCC 56

      56 rAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGlu..... 69
         ::::::     ||||||||:  :::  :::|||||
      57 CTCCATCAGCCACAGCTATTGG...ATTTCCCACCCAGAATCTTTAGCTA 103

      70 ..........AspSerSer......TrpAlaAlaLysAlaProGlyAla 80
                 |||||||||:      |||:::  :::          :::
     104 AATGAGATCAFGATTCTGGAAGGAGGTGGTGTAATGAA.....TCT 144

      81 SerSerArgGluGluProProGluGluProGlnGlnCysProValIleAs 97
         :::  |||:::::||||||
     145 CAACCCCGGCAACAACCTCCT.............................. 165

      97 pSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGlyLeuThrL 114
               ::|||||||:::|||||||||
     166 .......TCACCAGCCCCCAGCCTGGAC..................... 186

     114 euGlnGlu.HisSerLeuGluGlnValGlnSerMetValValGlyGluVa 130
```

```
         :::::: :::|||       :::||||:::|||          |||
     187 ..AGACAGCTACTCCACGTGCAATGTTCCAGTGGGTTTTTTGGAGGCCA 234

     130 lLeuLysAspIleGluArgAlaCysLysLeuLeuAsnIleThrAlaAspP 147
             :::|||||:                                   |
     235 GTGGCATGAAATTCAT................................C 251

     147 roMetAspTrpSerProSerAsnValGluLysTrpLeuLeuTrpThrGlu 163
         ||::::  |||:::       :::|||  :::|||||||
     252 CTCAGTACTGGACCAAGTACCAGGTGTGGGAGTGGCTCCAGCACCTCCTG 301

     164 HisGlnTyrArgLeuProProMetGlyLysAlaPheGlnGlu......Le 178
                 :::|||  ||||||||||       ::
     302 GACACCAACCAGCTGGATGCCAATTGTATCCCTTTCCAAGAGTTCGACAT 351

     178 uAlaGlyLysGlnLeuCysAlaMetSerGluGluGlnPheArgGluArgS 195
         :   |||:::::::|||||||:::|||||   :::::::|||   :::   :
     352 CAACGGCGAGCACCTCTGCAGCATGAGTTTGCAGGAGTTCACCGGGCGG 401

     195 erProLeuGlyGlyAspValLeuHisAlaHisLeu.............. 206
         ::      :::|||::::::::|||:::::::|||
     402 CAGGGACGAGGGGACGCTCCTCTACAGCAACTTGCAGCATCTGAAGTGG 451

     207 ..................AspIleTrpLysSer.........AlaAlaPr 214
                           |||:::::::::::|||       :::
     452 AACGGCCACTGCAGTAGTGACCTGTTCCAGTCCACACACAATGTCATTGT 501

     214 pMetLysGluArgThrSerProGlyAlaIle..................H 225
             |||:::|||:::|||:::   :::                    :
     502 CAAGACTGAACAAACTCAGCCTTCCATCATGAACACCTGGAAAGACGAGA 551

     225 isTyrCysAlaSerThrSerGluGlnSerTrpThrAspSerGluValAsp 241
         ::|||      :::|||:::      |||   :::|||      :::|||
     552 ACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTG...TTGGAC 598

     242 SerSer......CysSerGlyGln........................... 247
         |||:::      |||   :::|||
     599 AGCAAAACTTTCTGCCGGGCTCAGATCTCCATGACAACCACCAGTCACCT 648

     248 .Profile.......................................... 249
          ||||::
     649 TCCTGTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCCCCCTGCCA 698

     250 .....................................HisLeuTrpGlnPhe 254
                                            |||||||||:::|||
     699 AGTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGAATTC 748

     255 leuLysGlnLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTr 271
         ::::::::::::|||||:::|||   :::          :::|||:::||
     749 ATCCGGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATAAAATG 798

     271 pLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnValA 288
         |   :::::::::::::|||:::|||:::::::      |||      ||||
     799 GGAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGCAGTGG 848

     288 laArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnGlyAspLys 304
         |||:::|||||||||   :::|||||||:::   :::|||:::|||:|||
     849 CTCAGCTATGGGGTAAAAAGAAGAACAACAGCAGCATGACCTATGAAAAG 898

     305 LeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPr 321
         ||||||||||:::::|||   ||||||||||:::   |||:::::::::
     899 CTCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATTCTGGAGCGTGT 948

     321 oAspIleSerGluArgLeuValTyrGlnPhe 341
          |||   :::::|||||||!|||||:::|||
     949 GGAT...GGACGAAGACTGGTATATAAATTT 976

seq_name: qb_rol:AF035527

seq_documentation_block:
LOCUS       AF035527      1319 bp    mRNA            ROD       19 MAY 1998
DEFINITION  Mus musculus EHF (Ehf) mRNA, complete cds.
ACCESSION   AF035527
```

```
VERSION     AF035527.1  GI:3138929
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1319)
  AUTHORS   Bochert,M.A., Kleinbaum,L.A., Sun,L.Y. and Burton,F.H.
  TITLE     Molecular cloning and expression of Ehf, a new member of the ets
            transcription factor/oncoprotein gene family
  JOURNAL   Biochem. Biophys. Res. Commun. 246 (1), 176-181 (1998)
  MEDLINE   98262938
REFERENCE   2  (bases 1 to 1319)
  AUTHORS   Bochert,M.A., Kleinbaum,L.A., Sun,L.Y. and Burton,F.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-1997) Pharmacology, University of Minnesota, 425
            Delaware St. SE, 8-249 Millard Hall, Minneapolis, MN 55455, USA
FEATURES             Location/Qualifiers
     source          1. .1319
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="pituitary"
     gene            1. .1319
                     /gene "Ehf"
     CDS             144. .1046
                     /gene "Ehf"
                     /codon_start 1
                     /product "EHF"
                     /protein_id "AAC40119.1"
                     /db_xref="GI:3138930"
                     /translation="MILEGSGVMNLNPANNLLHQQPAWTDSYPICNVSSGFFGSQWHE
                     IHPQYWTKYQVWFWLQHLLDLNQLDASCIPFQEFDISGPHLCSMSLQEFIRAACSACQ
                     LLYSNLQHLKWNGQCSSGLEQSAHNVIVKTEQTDESIMNIWKEENYLYDPSYGSTVDL
                     LDSKIFCRAQISMTTSSHLPVAESPDMKKFQDHPVKSHTKKHNPFGIHLWEFIRDILL
                     SPDKNPGLIFWEDRSEGTFPGLLESEAVAQIWGEEKNNSSMGYEDLSEAMRYYYERELL
                     ERVDGRRLVYKPGKNARGWPFNFN"
BASE COUNT       397 a     309 c     299 g     314 t
ORIGIN


alignment_scores:
            Quality:  256.50              Length:     246
              Ratio:    1.914               Gaps:       9
 Percent Similarity:   54.472   Percent Identity: 28.049

alignment_block:
 US-09-126-945B-2 x AF035527   ..

 Align seq 1/1  to: AF035527  from: 1  to: 1319

      147 ProMetAspIrpSerProSerAsnValGlnLysIrpLeuLeuIrpHarGl 163
          ||:::  |||:::      :::|||  :::||||||
      282 CCICAGTACTGGACCAAATACCAGGTGTGGGAATGGCTGCAGCACCTGCT 331

      163 uHisGlnTyrArgLeuProProMetGlyLysAlaPheGluGlu.....L 179
               :::|||              |||||||||          :
      332 GGACACCAACCAGCTAGACGCTAGCTGCATCCCTTTCCAGGAGTTCGACA 381

      178 euAlaGlyLysGluLeuCysAlaMetSerGluGluGlnPheArgGlnArg 194
          :::::|||:::::|||||::||||    :::::::||    :::
      382 TTAGCGGAGAACACCTGTGCAGCATGAGTCTGCAGGAGTTCACGAGGGCA 431

      195 SerProLeuGlyGlyAspValLeuHisAlaHisLeu                228
          :::      :::|||:::::::::::::::||||
      432 GCAGGCTCAGCTGGGCAGCTGCTCTACAGCAACCTACAGCATCTCAAGTG 481

      207 ..................AspIleTrpLysSerAla.........Alal 214
                           |||:::::::::|||||
      482 GAACGGCCAATGCAGCAGTGACCTTTCCAGTCCGCACACAATGTCATTG 531

      214 rpMetLysGluArgThrSerProGlyAlaIleHisTyrCysAlaSerThr 240
             ||||:|||:::|||:::   ::::::          :::
      532 TCAAGACTGAACAAACCGATCCTTCCATCATGAAC.........ACATGG 574

      231 SerGluGluSerTrpThr...............AspSerGluValAsp... 241
          :::|||||||:::::                  |||   ||||||
      575 AAAGAAGAAAACTATCTCTATGATCCCAGCTATGGTAGCACAGTAGATCT 622

      242 ..............SerSerCysSerGlyGln..................... 247
                        :::   |||   :::|||
      623 GTTGGACAGTAAGACTTTCTGCCGGGCTCAGATCTCCATGACAACCTCCA 672

      247 ..................................................... 247

      673 GTCACCTTCCAGTTGCAGAGTCACCTGATATGAAAAAGGAGCAAGACCAC 722

      248 ProHe.....................................HisLeuTr 252
          |||:::                                  ||||||||
      723 CCTGTAAAGTCCCACACCAAAAACACAACCCAAGACGCACTCACTTATG 772

      252 pGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPhel 269
          |:::|||:::::::::::::|||||||::|||   :::         :::|
      773 GGAGTTCATCCGAGACATTCTCTTGAGCCCACACAAGAACCCAGGACTGA 822

      269 leArgIrpLeuAsnLysGluLysGlyLeuHisLeuLysLeuGluAspSerAla 285
          |||:::|||   :::::::::::|||||||||:::::       |||
      823 TCAAATGGGAAGACCGTTCGGAAGCCATCTTCAGGTTCCTGAAGTCAGAA 872

      286 GlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTy 302
          |||:|||:::|||||||||   :::|||||||:::   :::|||:::||
      873 GCTGTGGCTCAGCTGTGGGGGAAAAAGAAAAATAACAGTAGCATGACATA 922

      302 rAspLysLeuSerArgSerIleArgGluTyrTyrLysLysGlyIleIleA 319
          |:::|||||||||||||:::::|||   |||||||||:::   |||:::
      923 GGAGAAGCTCAGCCGGGCTATGAGATATTACTACAAAGGAGAAATCCTGG 972

      319 rgLysProAspIleSerGlnArgLeuValTyrGlnPhe 331
          :::::   |||   ::::::|||g|||||||:|||
      973 AACGTGTGGAT   GGACGACGGCTGGTCTATAAGTTT 1007

seq_name: gb_pr5:AF124439

seq_documentation_block:
LOCUS       AF124439     2586 bp    mRNA            PRI       27-JAN-2000
DEFINITION  Homo sapiens transcription factor PSE-3B (PSE-3) mRNA, complete
            cds.
ACCESSION   AF124439
VERSION     AF124439.1  GI:4838433
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2586)
  AUTHORS   Kas,K., Finger,E., Grall,F., Gu,X., Akbarali,Y., Bottaro,J.,
            Weiss,A., Oettgen,P., Kapeller,R. and Libermann,T.A.
  TITLE     ESE-3, a novel member of an epithelium specific ets transcription
            factor subfamily, demonstrates different target gene specificity
            from ESE-1
  JOURNAL   J. Biol. Chem. 275 (4), 2986-2998 (2000)
  MEDLINE   20112875
REFERENCE   2  (bases 1 to 2586)
  AUTHORS   Kas,K., Finger,E., Grall,F., Gu,X., Akbarali,Y., Bottaro,J.,
            Oettgen,P., Kapeller,R. and Libermann,T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-1999) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1. .2586
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "11"
                     /map="11p14"
                     /tissue_type="prostate gland"
     gene            1. .2586
                     /gene="ESE-3"
```

```
     CDS             98. .1000
                     /gene "ESE 3"
                     /note "Ets transcription factor"
                     /codon_start 1
                     /product "transcription factor ESE-3B"
                     /protein_id "AAD30991.1"
                     /db_xref "GI:4884844"
                     /translation "MILEGGGVMNLNPGNNLLHQPPAWTDSYSICNVSSGFFCSQWHE
                     IHPQYWTKYQVWEWLQHLLDTNQLDANCIPPQEFDINGEHLCSMSLQEFTKAAGTAGQ
                     LLLYSNLQHLAWNSQESSSDLFQSTHNVLVKTEQTEPSTMNTWKDENYLYDLNYCSTVDL
                     LDGKIFCKVQISMTLTSHLPVAESPDMKEEQDPPAPCHEKEHNDPCPHLWEFTPTLLL
                     NPDENPGLLEWEDPSEGVFEKFLESFAVAQGWCEFENN-SMEYFFISPAMFTYFRETL
                     ERVDGRRLVYKEGKNARGWRENEN"
BASE COUNT       793 a    543 c    573 g    687 t
ORIGIN


alignment scores:
            Quality:  256.00              Length:     360
              Ratio:    1.524               Gaps:      10
 Percent Similarity:   46.667   Percent Identity:  23.333

alignment block:
 US-09-126-945B-2 x AF124439   ..

 Align seq 1/1  to: AF124439  from: 1  to: 2586

       38 LeuGluArgArgAspTrpSerProSerProProAlaThrProGluGlnGl  54
          |||   |||   :::   :::   ||||||||||::|||          ||
       14 CTGGTGAGATTGCAGCCTGCTGCCTCCCCTCCATCAACCACAGCTATTGG  63

       54 yLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspS  71
          |:::   :::       |||         :::::::::      |||   :
       64 ATTTCCCACCCAGAATCTTTAGCTAAATGAGATCATGATTCTGGAAGGAG  113

       71 erSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluProPro  87
          :::::                |||||   :::      ::::::||||||
      114 GTGGTGTAATGAATCTCAACCCCGGCAACAACCTCCTTCACCAGCCGCCA  163

       88 GluGluProGluGlnCysProValIleAspSerGlnAlaProAlaGlySe  104

      164 ..................................................  168

      104 rLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGluG  121
                                     ::::::::::|||      :
      164 .....................GCCTGGACAGACAGCTACTCCACATGCA  191

      121 lnValGlnSerMetValValGlyGluValLeuLysAspIleGluThrAla  137
          ::|||:::|||         ||:::         :::|||:::
      192 ATGTTTCCAGCGGTTTTTTTGGAAGCCAGTGGCATGAAATCCAC...... 235

      138 CysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAs  154
                                  |||:::   |||:::      ::
      236 ........................CCTCAGTACTGGACCAAGTACCA  258

      154 nValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProM  171
          :|||   :::|||||                    :::|||
      259 GGTGTGGGAGTGCCTCCAGCACCTCCTGGACACCAACCAGCTGGATGCCA  308

      171 etGlyLysAlaPheGlnGln......LeuAlaGlyLysGluLeuCysAla  185
                |||||||||      :::   |||::::::::::::::::
      309 ATTCTATCCCTTTCCAAGAGTTCGACATCAACGGCGAGCACCTTTGCAGC  358

      186 MetSerGluGluGluPheArgGlnArgSerProLeuGlyGlyAspValLe  202
          ||||||   :::::|||   :::   :::      :::|||::::::||
      359 ATGAGTTTGCAGGAGTTCACCCGGCGGCAGGGACGGCGGGGCAGCTCCT  408

      202 uHisAlaHisLeu.................................Asp1  208
          |::::::::|||                                  |||:
      409 CTACAGCAACTTGCAGCATCTGAAGTGGAACGGCCAGTGCAGTAGTGACC  458

      208 IeTrpLysSer.........AlaAlaTrpMetLysGluArgThrSerPro  221
          :::::::|||         :::          ||||:|||::|||

      459 TGTTCCAGTCCACACACAATGTCATTGTCAAGACTGAACAAACTGAGCCT  508

      222 GlyAlaIleHisTyrCysAlaSerThrSerGluGluSerIrpThrAsp..  237
          :::   ::::::                              ::::||   |||
      509 TCCATCATGAAC.......................ACCTGGAAAGACGA  544

      238 ..........................SerGluValAsp.............  241
                                     |||   ||||||
      545 GAACTATTTATATGACACCAACTATGGTAGCACAGTAGATTTGTTGGACG  594

      242 ..............................SerSerCysSerGlyGln  247
                                       ::::::   |||
      595 GCAAAACTTTCTGCCGGGTTGAGATCTCCATGACAACCACCAGTCACCTT  644

      248 Profile.....................................................  249
          ||||::
      645 CCTGTTGCAGAGTCACCTGATATGAAAAAAGAGCAAGACCCCCCTGCCAA  694

      250 ..................................HisLeuTrpGlnPheL  255
                                           ||||||||||::|||:
      695 GTGCCACACCAAAAAGCACAACCCGAGAGGGACTCACTTATGGGAATTCA  744

      255 euLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIleArgTrp  271
          ::::::::::|||||:::|||   :::         :::|||:::|||
      745 TCCGCGACATCCTCTTGAACCCAGACAAGAACCCAGGATTAATAAAATGG  794

      272 LeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGlnValAl  288
          :::::::::::::|||:::|||::::::      |||     |||||
      795 GAAGACCGATCTGAGGGCGTCTTCAGGTTCTTGAAATCAGAGGCAGTGGC  844

      288 aArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrAspLysL  305
          |::::::::::::   :::::::::::   ::::|::::|||||
      845 TCAGCTATGGGCTAAAAAGAAGAACAACAGCAGCATGACCTATGAAAAGC  894

      305 euSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArgLysPro  321
          ||||||||:::::::||   ||||||||||:::   |||:::::::::
      895 TCAGCCGAGCTATGAGATATTACTACAAAAGAGAAATTCTGGAGCGTGTG  944

      322 AspIleSerGluArgLeuValTyrGlnPhe  341
          |||   ::::::|||||||||||||||:||
      945 GAT...GGACGAAGACTGGTATATAAATTT  961

seq_name: gb_pr5:AF170583

seq_documentation_block:
LOCUS       AF170583      2590 bp    mRNA            PRI       10-NOV-1999
DEFINITION  Homo sapiens ets homologous factor (EHF) mRNA, complete cds.
ACCESSION   AF170583
VERSION     AF170583.1  GI:6318677
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2590)
  AUTHORS   Kleinbaum,L.A., Duggan,C., Ferreira,E., Coffey,G.P., Buttice,G. and
            Burton,F.H.
  TITLE     Human chromosomal localization, tissue/tumor expression, and
            regulatory function of the ets family gene EHF
  JOURNAL   Biochem. Biophys. Res. Commun. 264 (1), 119-126 (1999)
  MEDLINE   99458631
   PUBMED   10527851
REFERENCE   2  (bases 1 to 2590)
  AUTHORS   Kleinbaum,L.A., Coffey,G.P. and Burton,F.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-1999) Pharmacology, University of Minnesota,
            3-249 Millard Hall, 435 Delaware Street S.E., Minneapolis, MN
            55455, USA
FEATURES             Location/Qualifiers
     source          1. .2590
                     /organism="Homo sapiens"
                     /db_xref ":taxon:9606"
                     /chromosome "11"
```

```
                     /map="11p12"
                     /tissue_type "prostate"
     gene            1. .2590
                     /gene="EHF"
     CDS             102. .1004
                     /gene="EHF"
                     /codon_start 1
                     /product "ets homologous factor"
                     /protein_id "AAF06998.1"
                     /db_xref "GI:6318678"
                     /translation "MILEGGGVMNLNPGNNLLHQPPAWIDGYSIQNVHSSMTPSSQWHE
                     IHPQYWTKYQVWEWLQHLLDENQLDANCIPPEQEFDINGEHLCSMSLQEFTKAACTAGQ
                     LLYSNLQPLKWNGQGSSPLTPQSTHNVIVKTDVEFLESSIMNIWEIENYLYSINYGSIVDL
                     LDSSTFCKAQISMTTISHLPVAESPDMEEEQQPTAECHTKRNPSCIHLWEITEMIEL
                     NPSKNCSLEEWFVSFCVELLSCAVACLWCPEFNSCMPSSFSAMEYYYRHIL
                     ERVDGRRLVYKFGKNARGWRENEN"
BASE COUNT       792 a    534 c    578 g    686 t
ORIGIN


Alignment scores:
            Quality:  256.00              Length:     458
              Ratio:  1.488                 Gaps:      11
 Percent Similarity:  48.725   Percent Identity:  24.463

alignment block:
 US-09-126-945b-2 x AF170583  ..

 Align seq 1/1  to: AF170583  from: 1  to: 2590

      38 LeuGluArgArgAspTrpSerProSerProProAlaThrProGluGlnGl 54
         |||  |||  :::  :::  ||||||||||:::|||        ||
      18 CTGGTGAGATTGCAGCCTGCTGCCTCCCCTCCATCAACCACAGCTATTGG 67

      54 yLeuSerAlaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspS 71
         |:::  :::      |||        :::::::::      |||  :
      68 ATTTCCCACCCAGAATCTTTAGGTAAATGAGATCATGATTCTGGAAGGAG 117

      71 erSerTrpAlaAlaLysAlaProGlyAlaSerSerArgGluGluProPro 87
         :::::              ||||||  :::      ::::::::::::
     118 GTGGTGTAATGAATCTCAACCCCGGCAACAACCTCCTTCACCAGCCGCCA 167

      88 GluGluProGluGlnCysProValIleAspSerGlnAlaProAlaGlySe 104

     167 .................................................. 167

     104 rLeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGlnC 121
                                  :::::::::|||      :
     168 ......................GCCTGGACAGACAGCTACTCCACGTGCA 195

     121 lnValGluSerMetValValGlyGluValLeuLysAspIleGluThrAla 137
         ::|||:::|||         |||            ::::::::
     196 ATGTTCCAGTGGGTTTTTTGGAGGCCAGTGGCATGAAATTCAT...... 239

     138 CysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAs 154
                              ||:::  |||:::      ::
     240 ..........................CCTCAGTACTGGACCAAGTACCA 262

     154 nValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProM 171
         :|||  :::||||||          :::|||
     263 GGTCTGGGAGTGGCTCCAGCACCTCCTGGACACCAACCAGCTGGATGCCA 312

     171 etGlyLysAlaPheGlnGlu      LeuAlaGlyLysGluLeuCysAla 185
                 |||||||||       :::   |||:::::::::::::
     313 ATTGTATCCCTTTCCAAGAGTTCGACATCAACGGCGAGCACCTTTGCAGC 362

     186 MetSerGluGluGlnPheArgGlnArgSerProLeuGlyGlyAspValIle 202
         ||||||   ::::::::||   :::   :::      ::::|||:::::||
     363 ATGAGTTTGCAGGAGTTCACCCGGGCGGCAGGGACGGCGGGCCAGCTCCT 412

     202 uHisAlaHisLeu.................................AspI 218
         |::::::::::|||                                |||:
     413 CTACAGCAACTTGCAGCATCTGAAGTGGAACGGCCAGTGCAGTAGTGACC 462

     208 leTrpLysSer.........AlaAlaTrpMetLysGlnArgThrSerPro 221
         :::::::::|||         :::             |||:::|||:::|||
     463 TGTTCCAGTCCACACACAATGTCATTGTCAAGACTGAACAAACTGAGCCT 512

     222 GlyAlaIle.................HisTyrCysAlaSerThrSerGl 232
         :::  :::                 :::|||       :::|||:::
     513 TCCATCATGAACACCTGGAAAGACGAGAACTATTTATATGACACCAACTA 562

     232 qElaSerTrpThrAspSerGluValAspSerSer.....CysSerGlyG 247
           |||  :::|||      :::||||||:::      |||  :::|
     563 TGGTAGCACAGTAGATTTG...TTGGACAGCAAAACTTTCTGCCGGGGCC 609

     247 In.........................ProIle.............. 249
         ||                        ||:::
     610 AGATCTCCATGACAACCACCAGTCACCTTCCTGTTGCAGAGTCACCTGAT 659

     249 .................................................. 249

     660 ATGAAAAAGGAGCAAGACCCCCCTGCCGAGTCCCACACCAAAAAGCACAA 709

     250 .............HisLeuTrpGluPheLeuLysGluLeuLeuLeuLysP 262
                     ||||||||||:::|:::::::::::::||||||:::|
     710 CCCGAGAGGGACTCACTTATGGGAATTCATCCGCGACATCCTCTTGAACC 759

     262 roHisSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIle 278
         ||  :::          :::|||:::|||  ::::::::::::::|||:::
     760 CAGACAAGAACCCAGGATTAATAAAATGGGAAGACCGATCTGAGGGCGTC 809

     279 PheLysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLy 295
         |||:::::       |||      |||||:::::|||||||   :::||
     810 TTCAGGTTCTTGAAATCAGAGGCAGTGGCTCAGCTATGGGGTAAAAAGAA 859

     295 sAsnArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnT 312
         ||||:::   :::|||:::|||:::|||||||||||||:::::|||   |
     860 GAACAACAGCAGCATGACCTATGAAAAGCTCAGCCGAGCTATGAGATATT 909

     312 yrTyrLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuVal 328
         ||||||||:::   |||:::::::::   |||   :::::|||||||||
     910 ACTACAAAAGAGAAATTCTGGAGCGTGTGGAT...GGACGAAGACTGGTA 956

     329 TyrGlnPhe 331
         |||:::|||
     957 TATAAATTT 965

seq_name: qb_ins:DM74E

seq_documentation_block:
LOCUS        DM74E          4800 bp    DNA             INV       12 SEP 1993
DEFINITION   Drosophila melanogaster 74E gene.
ACCESSION    X15087
VERSION      X15087.1  GI:7513
KEYWORDS     74E gene.
SOURCE       fruit fly.
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 4800)
  AUTHORS    Ponqs,O.
  TITLE      Direct Submission
  JOURNAL    Submitted (26-APR-1989) Ponqs O., Lehrstuhl fuer Biochemie,
             Abteilung Chemie, Ruhr Universitaet Bochum, Universitaetsstr 150,
             D-4630 Bochum, F R G
REFERENCE    2  (bases 1 to 4800)
  AUTHORS    Jankuecht,R., Taube,W., Ludecke,H.J. and Ponqs,O.
  TITLE      Characterization of a putative transcription factor gene expressed
             in the 20-OH-ecdysone inducible puff 74EF in Drosophila
             melanogaster
  JOURNAL    Nucleic Acids Res. 17 (12), 4455-4464 (1989)
  MEDLINE    89315191
COMMENT      Data kindly reviewed (10-Jul-1989) by Ponqs O.
FEATURES             Location/Qualifiers
```

```
     source          1. .4800
                     /organism "Drosophila melanogaster"
                     /db_xref "taxon:7227"
                     /clone "lof4, lof5, 74E11, 74E12"
                     /map "chromosome 3"
     misc_feature    1. .700
                     /note "genomic sequence"
     CDS             534. .4185
                     /note "74E protein (AA 1 883)"
                     /codon_start 1
                     /protein_id "CAA33195.1"
                     /db_xref "GI:7514"
                     /db_xref "FLYBASE:FBgn0000567"
                     /db_xref "SWISS-PROT:P11536"
                     /translation "MEMVQHELVAASAHNPASQAAASLVNVSSSSSSSSSSSSSSLSLS
                     SSSSSSSLSSATPTPVASPVTPTSPPTPAAAAPAPASPPACAELQELQQQAETQEDPTM
                     KDQDMLEKTRQEVKDPVNVEEPGAIVDPESVMARQSPSPVASTKVIPESLEETSNKSPT
                     VQEDEEEESPSVASDCREFKVIYNHIRQQQHHHSPSSPDQKTRSLTDVSELTWEPKQQL
                     QRSSVITAAPTLQPQQHQQPMSDIEDEETLEDVDDADADVEADAEDEELLFQYQNGYD
                     SPLDLSIGGATSAAASAAAAASAVSPPPCPTYSCIESPDSAQCEPAAPMPLKPEPKAEE
                     SAAYKESLMKEYYTETPIVEQSISPAPQQQLQQQHHLQQQQQQQPHNGSTPAGATALL
                     HIKTEQNTLLTPLQLQQQQQQQQLHCAACNCQSSNENNAHQQQQPLATPQRPLLHNL
                     LSGGATHNPHHRNYTTATTGSFPPSPADSGVSDVDSSSSSGGQPCADELKARLGMPPAT
                     SASAAAAAAAAAAAAAAHLHTGTPLHPNLYQNNAANSLRNTWNRSVGVPDNYYGSSGAG
                     SGGTQPGGPGNPQTIPGYLTTSYFNAPTAATAAASQRGTLNGYHSLHQQQQQQQQSQQ
                     SQQQQQLAHQQLSHQQQQALHQQLSHQQQQQQQQQHPHSQLNGPHPHSHPHSHPHS
                     HPHAGQHTHSTLAAAAAAAAAASVVSSSSSAVAAAAMLSASAAAAATAAAAAGGSQSV1
                     QPATSSVSYDLSYMLELGGPQQRKAKKPPKKQKLEMGVKRESREGSTEYLWEELLELLQ
                     DREYCPREIPWTNPFKCVFELVDSKAVSMLWGMHELNEPDMNYETMGRALRYYYQRGIL
                     AKVDGQRLVYHFVDVPKDLIIETDCNGV"
     misc_feature    701. .4800
                     /note "cDNA sequence"
     polyA_site      4800
                     /note "polyA site"
BASE COUNT      1681 a    1276 c    1142 g    1001 t
ORIGIN


alignment scores:
            Quality:  256.00              Length:     385
              Ratio:  1.280                 Gaps:      19
 Percent Similarity:  51.948   Percent Identity: 27.792

alignment block:
 US-09-126-945B-2 x DM74E   ..

 Align seq 1/1  to: DM74E  from: 1  to: 4800

       2 GlySerAlaSerProGlyLeuSerSerValSerProSerHisLeuLeuLe 18
         ||||||:::::|||:::   :::   :::|||   |||:::::::::::
    2096 GGCAGCGGCACACCTGCACACGGGCACCTTTCCTGCATCCGAATTTATA.. 2143

      18 uProProAspThrValSerArgThrGly.....LeuGluLysAlaA 32
           |||:::     ::::::   :::          :::|||:::   :
    2144 ....CCAGAACAATGCGGCAAATTCGCTGCGGAACATATGGAATCGCAGT 2189

      32 luAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSer..... 46
         ::   ||||||      |||   |||::::::|||          :::
    2190 GTTGGGGTGCCCGACAACTACTACGGAAGCAGTGGCGCCGGCAGTGGGGG 2239

      47 ...Pro.ProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeu.... 60
           ||| |||:::   |||   :::       :::       ||||||
    2240 CACCCAGCCAGGTGGTCCGGGAAACCCCCAAACGCCTGGCTACCTGACAA 2289

      61 ..SerTyrPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLys 76
           |||||||||:::
    2290 CGTCCTACTTCAAC..................................... 2303

      77 AlaProGlyAlaSerSer.ArgGluGluProProGluGluProGluGlnC 93
         ||||||   |||:::::: ::::::::|||   ||||||||||:::::
    2304 GCACCGACAGCGGCAACGGCAGCGGCGTCCCAGCGAGGAACCACAATCAA 2353

      93 ysProValIle.AspSerGlnAlaProAlaGlySerLeuAspLeuValPr 109
```

```
                ::::|||   :::   |||   |||::::::      ::::::
    2354 TGGCTATCATCCCTGCACCAGCAGCAGCAGCAGCAACAGCAGTCGCAGC 2403

     109 oGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetV 126
                 :::   :::                          |||      :
    2404 AATCGCAGCAGCAACAGCAACTGGCCCACCAGCAACTGTCGCACCAGCAG 2453

     126 alValGlyGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsn 142
         :::::|||   :::       :::       ::::::      :::
    2454 CAGCAGGGCGTTGCATCAGCAACTGTCGCACCAACAGCAGCAGCAGCAACA 2503

     143 IleThrAlaAsp.............ProMetAspTrpSerProSerAsnVa 155
             ||||||                |||   :::   |||   |||   ::
    2504 GCAACAGCAGCAACAGCCGCACTGCCAACTGAACGGTCGCATCCGCACT 2553

     156 lGlnLysTrpLeuLeuTrpIleGlnHisGlnTyrArgLeuProProMetG 172
         :   :::   |||       :::       :::   |||   |||       :
    2554 CACATCCGCACTCTCATCCGCATTCGCATCCGCACGCCGCCCA.....G 2597

     172 lyLysAlaProGlnGluLeu.............AlaGlyLysGluLeuCys 184
         ::   ||||||::::::                 :::|||::::::::|||
    2598 CACACCCATTCCACAATCGCTCGCAGCGGCGCAGCGGCAGCACACATCTCT 2647

     185 AlaMetSerGluGluGlnPheArg.................GlnArgSe 195
             :::::::::::|||   |||                 :::|||::
    2648 CGTCAGCAGCAGCAGCAGCGCCGTTGCAGCGGAGCCATCGTGAGCGCCA 2697

     196 rProLeuGly..................................... 198
         :   |||
    2698 GTGCAGCGGCAGCAGCAACGGCGGCAGCTGCGGCGGCGGCATCCCAGAGT 2747

     199 ..........GlyAspVal.....................LeuHisAla 204
                   |||||||||                 |||||||||
    2748 GTTATCCAACGGCGACGTCCAGCGTCAGCTACGATCTCICCCACATGCT 2797

     205 HisLeuAspIleTrpLysSerAlaAlaTrpMetLysGlnArgThrSerPr 221

    2797 ..................................................... 2797

     221 oGlyAlaIleHisTyrCysAlaSerThrSerGluGla...SerTrpThrA 237
          ||||||   :::      |||:::   :::||||||   :::   |||:
    2798 .GGAGCTGGGCGGATTCCAGCAGCGGAAGGCGAAGAAGCCGCGCAAACCG 2846

     237 spSer.GlnValAsp.........SerSerCysSerGlyGlnProIleHi 250
         ::::: |||:::          |||   :::|||:::       ::
    2847 AAGCTGGAGATGGGCGTAAAGCGGGCGCAGCCGGGAGGGATCCACCACCTA 2896

     250 sLeuTrpGluPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyA 267
         :|||||:::|||||   :::|||||          :::          |
    2897 TCTGTGGGAGTTCCTCCTCAAACTGCTCCAGGATCGCGAATACTGTCCGC 2946

     267 rgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGlnAsp 283
         |||||||||:::||   |||:::||||||||:::|||||:::   |||
    2947 GTTTCATCAAGTGGCTGAACGGGCAGAAGGGCGTCTTCAAGCTGGTCGAC 2996

     284 SerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMe 300
         |||       |||:::||||||||||||:::|||||:::|||   ||
    2997 TCGAAGGCTGTGTCCGCCTGTGGGGCATCACAAGAACAAGCCCGGACAT 3046

     300 tAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyI 317
         |||||||:::   :::::||:::::::|||   ||||||:::::|||
    3047 GAACTACGACAAGTTGTCCCGCAGCATCCGCCAGTACTACAAGAAGGGCA 3096

     317 leIleArgLysProAspIleSerGluArgLeuValTyrGlnPheVaI 332
         ||:::   |||   |||   :::|||||||||||||||:::||||||
    3097 TACTGGCCAAGGTGGATCTGGCCAGCGGCTGGTCTACCATTCGTTC 3140

seq_name: gb_htg6:AC019512

seq documentation block:
LOCUS       AC019512    51405 bp    DNA             HTG       04 JAN 2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
```

```
ACCESSION   AC019512
VERSION     AC019512.1  GI:6665385
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 51405)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC 1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210037 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..51405
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    14383 a   10820 c   11265 g   14937 t
ORIGIN


alignment_scores:
            Quality:  255.50              Length:     382
              Ratio:  1.817                 Gaps:      19
 Percent Similarity:  59.785   Percent Identity:  28.272

alignment_block:
 US-09-126-945B-2 x AC019512/raw  ..

 Align seq 1/1  to reverse of: AC019512  from: 1  to: 51405

        1 MetGlySerAlaSerProGlyLeuSerSerValSerPro.........Se 14
          :::|||| |||||||          |||:::|||                 :: 
    25724 GTTGGTTCAGCAAGCCGCTTTTGTTCTCAAGTACTCATTTATATATTTGA 25675

       14 rHisLeuLeuLeuProProAspThrValSerArgThrGlyLeuGluLysA 41
          :   :::|||      ||||||:::                        :::: 
    25674 CTTTGTTCTTGCAGTGCCCGACAACTAC............TACGGAAGCA 25637

       31 laAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPro 47
          :::::|||||||:   |||  :::                 ||||::: 
    25636 GTGGCGCCGGCAGTGGGGGCACCCAG..............CCAGGTGGT 25602

       48 ProAlaThrProGluGlnGlyLeuSerAlaPheTyrLeu......SerTy 62
          |||:::::|||||:          :::      ||||||      |||||
    25601 CCGGGAAACCCCCAA.........ACGCCTGGCTACCTGACAACGTCCTA 25561

       62 rPheAspMetLeuTyrProGluAspSerSerTrpAlaAlaLysAlaProG 79
          ||||:::                              ||||||
    25560 CTTCAAC.....................................GCACCGA 25547

       79 lyAlaSerSer.ArgGluGluProProGluGluProGluGlnCysProVa 95
          |||:::::: ::::::::|||   |||||||||||:::::      :: 
    25546 CAGCGGCAACCGCAGCGGCGTCCCAGCGAGCAACCACAATCAATGGCTAT 25497

       95 Ille.AspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGly 111
          :|||   :::   |||   |||::::::       ::::::    :::  
    25496 CATTCCCTGCACCAGCAGCAGCAGCAGCAACAGCAGTCGCAGCAATCGCA 25447

      112 LeuThrLeuGluGluHisSerLeuGluGlnValGlnSerMetValValGl 128
          :::                      |||                 ::::::|| 
    25446 GCAGCAACAGCAACTGGCCCACCAGCAACTGTCGCACCAGCAGCAGCACC 25397

      128 yGluValLeuLysAspIleGluThrAlaCysLysLeuLeuAsnIleThrA 145
          |   :::         :::      ::::::         :::   |||| 
    25396 CGTTGCATCAGCAACTGTCGCACCAACAGCAGCAGCAGCAACAGCAACAG 25347

      145 laAsp.............ProMetAspIrpSerProSerAsnValGluLys 157
          ||                  |||   :::   |||  |||   :::   ::: 
    25346 CAGCAACAGCGGCACTCCCAACTGAACGGTCCGCATCCGACTCACATCC 25297

      158 IrpLeuLeuTrpThrGluHisGluTyrArgLeuProProMetGlyLysAl 174
          |||      :::        :::   |||  |||      :::   || 
    25296 GCACTCTCATCCGCATTCGCATCCGCACGGCGGGCCA......GCACACGC 25253

      174 aPheGlnGluLeu.............AlaGlyLysGluLeuCysAlaMetS 187
          ||||::::::                :::|||:::::::::::::|||   :::: 
    25252 ATTCCACAATCGCTGCAGCGGCGGCAGCGGCAGCAGCATCTGTCGTCAGC 25203

      187 erGluGluGlnPheArg.................GlnArgSerProLeu 197
          ::::::::::|||   |||                 :::|||::: 
    25202 AGCAGCAGCAGCGCTGTCGCAGCGGCAGCCATGCTGAGCGCCCAGTGCAGC 25153

      198 Gly..................................................... 198
          |||
    25152 GGCAGCAGCAACGGCGGCAGCTGCGGCGGGCGGATCCCAGAGTGTTATCC 25103

      199 ....GlyAspVal.....................LeuHisAlaHisLeuA 207
              |||||||||                   |||||||| 
    25102 AACGGAGACGTCCAGCGTCAGCTACGATCTCTCCTACATGCT....... 25060

      207 spIleTrpLysSerAlaAlaIrpMetLysGluArgThrSerProGlyAla 223
                                                    |||||| 
    25059 ..................................................GGAGCT 25054

      224 IleHisTyrCysAlaSerThrSerGlnCls...SerTrpThrAspSer.G 239
          :::     ||:::   :::|||||   :::   |||::::::.| 
    25053 CGGCGCATTCCACCACGCGAAGCTCAACAACCCGCCCAAACCCAACCTGCC 25004

      239 lnValAsp.........SerSerCysSerGlyGlnProIleHisLeuTrp 252
          ||:::            |||  :::|||:::       :::||||| 
    25003 AGATGGGCGTAAAGCGGCGCAGCCGGAGGGATCCACCACCTATCTGTGG 24954

      253 GlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIl 269
          :::|||||   :::|||||          :::          |||||||| 
    24953 CAGTTCCTCCTCAAACTGCTCCAGGATCGCGAATACTGTCGCGTTTCAT 24904

      269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaG 286
          |:::|||   |||:::|||||||||:::||||||:::   ||||||
    24903 CAAGTGGACGAACGGGAGAAGGGCGTCTTCAAGCTGGTCGACTCGAAGG 24854

      286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
          |||:::|||||||||||:::::|||||:::|||   |||||||||
    24853 CTGTGTCCCGCCTGTGGGGCATGCACAAGAACAAGCCGGACATGAACTAC 24804

      303 AspLysLeuSerArgSerIleArgGlntyrTyrLysLysGlyIleIleAr 319
          :::   ::::::|||:::::::|||   |||||:::::::|||||:::
    24803 GAGACGATGGGCAGGGCGCTGAGGTACTACTACCAGCGCGGCATACTGGC 24754

      319 glysProAspIleSerGlnArgLeuValTyrGluPheVal 332
          |||   |||   :::|||||||||||||||||||||||||| 
    24753 CAAGGTGGAT...GGCCAGCGGCTCGTCTACGAGTTCGTG 24717

seq_name: gb_in1:AE003523

seq_documentation_block:
LOCUS       AE003523   283337 bp    DNA             INV       06-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386750 Section 45
            of 54, complete sequence.
ACCESSION   AE003523 AE002602
VERSION     AE003523.2  GI:10727853
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 283337)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
```

```
                George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
                Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
                Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
                Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
                Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
                Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
                Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
                Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
                Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
                Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
                Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
                Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
                Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
                Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
                Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
                Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
                Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
                Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
                Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
                Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
                Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
                Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
                McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
                Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
                Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
                Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
                Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
                Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Sidén-Kiamos,I.,
                Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
                Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
                Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
                Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
                Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
                Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
                Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
                Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 28333)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7293946.
FEATURES             Location/Qualifiers
     source          1..28333
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     mRNA            join(13235..14917,15059..15422,15757..15941,16004..16147,
                     16212..16388,16527..17309)
                     /gene="CG7571"
                     /product="CT23161"
                     /db_xref="FLYBASE:FBan0007571"
                     /db_xref="FLYBASE:FBgn0036732"
     gene            13235..17306
                     /gene="CG7571"
                     /map="74d1-74d1"
                     /db_xref="FLYBASE:FBan0007571"
                     /db_xref="FLYBASE:FBgn0036732"
     CDS             join(13510..14917,15059..15422,15757..15941,16004..16147,
                     16212..16388,16527..16708)
                     /gene="CG7571"
                     /note="CG7571 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0007571"
                     /db_xref="FLYBASE:FBgn0036732"
                     /protein_id="AAF49332.2"
                     /db_xref="GI:10727854"
                     /translation="MEFSNGSVEAAAGVGSGSESNGSGNGSGSGSGSGSGSGSGSGS
                     AFTPLLSQHNNGTTNGEVTTPPPSTVLYESTPSNNNLWKAFEDLGHLKNGLGNTLSSN
                     NNGTGNGHSLSEKYAHEQAFLTGGYKLPTRSSESEESDFDSDLNGGSSAESSSSCGLF
mRNA            complement(17733..>19497)
                /gene="CG6422"
                /product="CT19778"
                /db_xref="FLYBASE:FBan0006422"
                /db_xref="FLYBASE:FBgn0036733"
gene            complement(<17733..>19497)
                /gene="CG6422"
                /map="74D1-74D2"
                /db_xref="FLYBASE:FBan0006422"
                /db_xref="FLYBASE:FBgn0036733"
CDS             complement(17733..19394)
                /gene="CG6422"
                /note="CG6422 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0006422"
                /db_xref="FLYBASE:FBgn0036733"
                /protein_id="AAF49331.1"
                /db_xref="GI:7293972"
                /translation="MSEDDDQQYDKEQQKTRVGSTEPSPKKRQNAAASGAAATTTSGI
                TASSGAGTTTSTGGGTESTTSTSTSTEDSTRKTSNAEQACAPPPTAATLANKTDDDYF
                DLEMEMERDKVALLEEFERKKRAPQTNVSTDDLTKSNDQLNEPTCYEGDPAERKR
                KRKLLAGLGLNATNKEQYTELLRKQQQKQQQAIWYHSGGGRTAKGWGADYGGR
                AEDRGVKAEEALEVPSAAPAGRMVEMQNKTQSTAPTCSQVGDTRPVSSAAFSEDSSTL
                LTSSWSGLCKLWSVPDCELKQTLRGHASYVGGVALRPGVKADEENVVAMASGGHDGAV
                KLWGFNNEFSTADTGHMPHRVSKVAFHPSGRFLATACYTSSWRLWDLEQNTEVLEQF
                GHAKPVHTLSYHSDGSVLVTGGLDAFCRVWDLRTGRCIMFLEGHLGAVFGVDFSPNGF
                HIATGSQDNTCKLWDLRKRQPVYTLTAHLNLTSDVKYQQDCGSELVTCSYDSTIKLWS
                NKIWQPLKILQGHSNKVISVLHAPNSQYIATASFDKTIKFLWSPDS"
mRNA            19912..>21443
                /gene="CG7564"
                /product="CT23137"
                /db_xref="FLYBASE:FBan0007564"
                /db_xref="FLYBASE:FBgn0036734"
gene            <19912..>21443
                /gene="CG7564"
                /map="74D1-74D2"
                /db_xref="FLYBASE:FBan0007564"
                /db_xref="FLYBASE:FBgn0036734"
CDS             20081..21443
                /gene="CG7564"
                /note="CG7564 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0007564"
                /db_xref="FLYBASE:FBgn0036734"
                /protein_id="AAF49330.1"
                /db_xref="GI:7293971"
                /translation="MSAPSNKMSATPQMMAMETQLMGLLRKNEDERQLKESDPKVCKSF
                LLDCCPHDLLASIRMDLGECPKVHDLAFRADYESAAKIRDYYYDIEAMEHLQAFTADC
                DEFIISAKQFLKSLIQLELLAEVAEKANAVIGLAELLGKKLAKALAGLAGLVLGSMEL
                MKKLLKKLKAKKIKAAMKYKISMPASIYQQQKLRVCEVCSAYDSTHDNGLPLALDPEGSK
                LHLCPLTIKEKLIELEKIAAPKKAELKKIGEMIDGEDAGKGKNKYFVGGRELDRKSKV
                HKSKSKSKSKSKNKPSPSKSKSKSNGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGS
                KSRDHRERERESGRRDRDRHGRNDRGRFGSRGGGGGGGGHHKDDRRKSKSKEKSKEKR
                NKNHKRKSGSGSGNGQRKRSYSKRKNYKR"
mRNA            complement(join(32522..32912,32968..33052,33115..33275,
                33338..34252,34541..35475))
                /gene="CG6411"
                /product="CT19742"
                /db_xref="FLYBASE:FBan0006411"
                /db_xref="FLYBASE:FBgn0036735"
gene            complement(<32522..>35475)
```

```
                     /gene "CG6311"
                     /map "74D1-74D2"
                     /db_xref="FLYBASE:FBan0006311"
                     /db_xref="FLYBASE:FBqn0036735"
     CDS             complement(join(32591..32912,32968..33072,33115..33275,
                     33338..34252,34541..35100))
                     /gene "CG6311"
                     /note "CG6311 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0006311"
                     /db_xref="FLYBASE:FBqn0036735"
                     /protein_id="AAF49329.1"
                     /db_xref="GI:7293970"
                     /translation="MGPTDQDWIGCAVSIACDEVLGVFGTLIKQISAEELLVRAPRN
                     GVPLRKQNAFVVLKGTLDLSTDLIEFAKQDEIGHIAGIVVKPLTSYTLRTCNTLEV
                     QQQTQTQQQQQQTQTQQQQKQQDQFQFPPDDLPSTPRSPANDNSRVAAGGASSSSGPR
                     GNFNSALSDKMHQLKLIELNGSNGLLDRIPQLSKASSALQQQQTSSLDNSVAALLGNMI
                     PPKVFVKLGSYVSNTPESYCSSSQDSGEATGLSLQSSKPTLDLVSNCGQFYKQIAASSY
                     GNTNGNVPNNGNANNNGNGTGNGYPNGNGNGNGENPTNEVPLLSSMPQQCVALLDSS
                     EADDPLIHEDFDFFGNLALEDKQAIWDQTESITQKPQVVSGLVNNHHLKPGLSYRHDE
                     NILASKPLQLRQIESMFGQSQDFVIDQGLIIPTIPATVPNSLLESAMKAGLSLQQQTP
                     ILARGASDLAITLLGGAREITPANNHQWPKIAIGGQGKNMEINTMAAFGKQASHG
                     LIVLLYVEQAKLLPLNSSSPFISLFFATQNVIVHSVDAFTPDQLVTLSINIANLSDAI
                     RKWLSVNRASVLAIDPPPQGINEVAIKYSILPILFLNGTISIAIFGSSSSAAAAIPTPIA
                     STSAAASAIKSAASLNNCCKLYGCNLCIRKCLYRQCSLKSECYSHEYVCPTHSPD"
     mRNA            join(35750..36054,36441..36667,44301..44383,44453..44534,
                     45095..45183,45206..45375,46330..46477,46541..46665,
                     46940..47042,47315..47605,47839..48023,48052..48274,
                     48530..48601,48644..48724,49403..49643,49699..49823,
                     49895..50036,50100..50256,50316..50609,50674..51072,
                     51147..>51249)
                     /gene "Nedd4"
                     /product "CG3113"
                     /db_xref="FLYBASE:FBan0007555"
                     /db_xref="FLYBASE:FBqn0036736"
     gene            <35750..>51249
                     /gene "Nedd4"
                     /note "CG7555"
                     /map="74D1-74D2"
                     /db_xref="FLYBASE:FBan0007555"

alignment_scores:
            Quality:  255.50              Length:     382
              Ratio:  1.817                 Gaps:      19
 Percent Similarity:  50.785   Percent Identity: 28.272

alignment_block:
 US-09-126-945B-2 x AE003523/rev  ..

 Align seq 1/1  to reverse of: AE003523  from: 1  to: 283337

       1 MetGlySerAlaSerProGlyLeuSerSerValSerPro.....Se 14
         :::|||||||||||||        |||:::|||           ::
   68462 GTTGGTTCAGCAAGCCGCTTTTGTTCTCAAGTACTCATTTATATATTTGA 68412

      14 rHisLeuLeuLeuProProAspThrValSerArgIhrGlyLeuGluLysA 31
         :   :::|-|      |||||||:::                 ::::
   68411 CTTTGTTCTTGCAGTGCCCGACAACTAC.........TACGGAAGCA 68374

      31 laAlaAlaGlyAlaValGlyLeuGluArgArgAspTrpSerProSerPro 47
         :::::|||||||:   |||  :::              ||||::
   68373 GTGGCGCCGGCAGTGGGGGCACCCAG.............CCAGGTGGT 68339

      48 ProAlaIhrProGluGlnGlyLeuSerAlaPheTyrLeu.....SerTy 62
         |||::::::|||::          :::      ||||||      |||||
   68338 CCGGGAAACCCCCAA.........ACGCCTGGCIACCTGACAACGTCCIA 68298

      62 rPheAspMetLeuTyrProGluAspSerSerIrpAlaAlaLysAlaProG 79
         |||||::                          ||||||
   68297 CTTCAAC.....................................GCACCGA 68284

      79 lyAlaSerSer.ArgGluGluProProGluGluProGluGlnCysProVa 95
           |||:::::: ::::::::::|||   |||||||||:::::      ::

   68284 CAGCGGCAACGGCAGCGGCGTCCCAGCGAGGAACCACAATCAATGGCTAT 68234

      95 lIle AspSerGlnAlaProAlaGlySerLeuAspLeuValProGlyGly 111
         ::||    :::   |||  |||::::::       ::::::      :::
   68233 CATTGCTGCACCAGCAGCAGCAGCAGCAACAGCAGTCGCAGCAATCGCA 68184

     112 leuthrLeuG rGluHisSerLeuGluGlnValGlnSerMetValValGl 128
         :::                          |||                 :::::||
   68183 GCAGCAACAGCAACTGGCCCACCAGCAACTGTGCAGCAGCAGCAGCAGG 68134

     128 yGluValLeuLysAspIleGluIhrAlaCysLysLeuLeuAsnIleIhrA 145
         |  :::        :::       ::::::          :::   ||||
   68133 CGTTGCATCAGCAACTGTCGCACCAACAGCAGCAGCAGCAACAGCAACAG 68084

     145 lnAsp          ProMetAspTrpSerProSerAsnValGlnLys 157
         ||              |||  :::  |||  |||  :::  :::
   68083 CAGCAACA[illegible]GCAACTGAACGATCGCGCATCGCACTCACATCC 68034

     159 TrpLeuLeuTrpIleGlnHisGlnTyrArgLeuProProMetGlyLysAl 174
           |||     :::      :::  |||  |||     :::   ||
   68033 GCACTCTCATCGGCATTCGCATCGCACGCGGCCA......GCACACGC 67990

     174 aPheGlnGluLeu.............AlaGlyLysGlnLeuCysAlaMetS 187
         ||||::::::                :::|||||||||||||||   ::::
   67989 ATCCACAACATCCTCCACGGCGCTACGCGCAGCATCATCGTCGTCAGC 67940

     187 erGluGluGlnPheArg.................GlnArgSerProLeu 197
         :::::::::|||   |||                :::|||:::
   67939 AGCAGCAGCAGCGCCGTTGCAGCGGCAGCCATGCTGAGCGCCAGTGCAGC 67890

     198 Gly..................................................  198
         |||
   67889 GGCAGCAGCAACGGCGGCAGCTGCGGCGGCGCGATCCCAGATGCTATCC 67840

     199 ....GlyAspVal.....................LeuHisAlaHisLeuA 207
             |||||||||                   ||||||||
   67839 AACGGGCGACGTCCAGCGTCAGCTACGATCTCTCCTACATGCT....... 67797

     207 spIleTrpTyrSerAlaAlaTrpMetLysGlnArgProSerProGlyAla 226
                                                 ||||||
   67796 ..................................................GGAGCC 67791

     224 IleHisTyrCysAlaSerIhrSerGluGlu...SerTrpThrAspSer.G 239
         :::      ||::::  ::|||||  :::  ||||::::: |
   67790 GGGCGGATCCAGCAGCGGAAGGCGAAGAAGCCGCGCAAACCGAAGCTGG 67741

     239 luValAsp.........SerSerCysSerGlyGlnProIleHisLeuTrp 252
         ||:::             |||  :::|||:::      :::||||||
   67740 AGATGGCGTAAAGCGGCGCAGCCGGCAGGATCCACCACCTATCTGTGG 67691

     252 GlnPheLeuLysGluLeuLeuLysProHisSerTyrGlyArgProGlu 269
         :::|||||  :::|||||       :::      |||||||
   67690 GAGTTCCTCCTCAAACTGCTCCAGGATCGCGAATACTGTCCGCGTTTCAT 67641

     269 eArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaG 286
         |:::|||  |||:::|||||||||:::|||||:::   ||||||
   67640 CAAGTGGACGAACCGGGAGAAGGCGTCTTCAAGCTGGTCGACTCGAAGG 67591

     286 lnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyr 302
           |||:::|||||||||||||:::::::|||||:::|||   |||||||||
   67590 CTGTGTCCCGCCTGTGGGGCATGCACAAGAACAAGCGGCGACATGAACTAC 67541

     303 AspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleAr 319
         :::   :::::|||:::::::|||   ||||||:::::::::::::::::
   67540 GAGACGATGGGCAGGGCGCTCAGGTACTACTACCAGCGCGGCATACTGGC 67491

     319 qLysProAspIleSerGlnArgLeuValTyrGlnPheVal 332
           |||   |||   :::|||||||||||||||||||||||
   67490 CAAGGTGGAT...GGCCAGCGGCTGGTCTACCAGTTCGTG 67454

seq_name: qb_pr10:HSU73843

seq_documentation_block:
```

```
LOCUS       HSU73843     1915 bp    mRNA            PRI       06 AUG 1997
DEFINITION  Homo sapiens epithelial specific transcription factor ESE 1b
            (ESE 1) mRNA, complete cds.
ACCESSION   U73843
VERSION     U73843.1  GI:1841522
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1915)
  AUTHORS   Oettgen,P., Alani,R.M., Barcinski,M.A., Brown,L., Akbarali,Y.,
            Boltax,J., Kunsch,C., Munger,K. and Libermann,T.A.
  TITLE     Isolation and characterization of a novel epithelium specific
            transcription factor, ESE 1, a member of the ets family
  JOURNAL   Mol. Cell. Biol. 17 (8), 4419-4433 (1997)
  MEDLINE   97378019
REFERENCE   2  (bases 1 to 1915)
  AUTHORS   Oettgen,P., Akbarali,Y., Barcinski,J., Boltax,J., Kunsch,C. and
            Libermann,T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-1996) Medicine, Beth Israel Hospital, 330
            Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1..1915
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="adult liver, pancreatic carcinoma"
                     /chromosome="1"
                     /map="1q42.1"
     gene            1..1915
                     /gene="ESE-1"
     CDS             120..1235
                     /gene="ESE-1"
                     /note="alternative splice form of ESE-1,
                     epithelial specific; member of the Ets transcription
                     factor family"
                     /codon_start=1
                     /product="ESE-1b"
                     /protein_id="AAB65823.1"
                     /db_xref="GI:1841523"
                     /translation="MAATCFTSNIFSNYFSAMYSSEDSTLASVPPAATFGADDLVLLL
                     SNPQMSLEGTEKASWLGEQPQPWSKTQVLDWLSYQVEENKYDASALTDFSRCDMDGATL
                     CNCALEELRLVPGPLGDQLHAQLRDLTSSSSDFELSWTFLLEKHCMAPQEALDEGPED
                     QGSPFAQELLDDGQQASPYHPGSCGAGADSPGSSDVSTAGTGASRSSHSSDSGGSDVD
                     LDPTDGKLEFSPWFEPCKGDPKHCKKFCGKPFKLSKGYWPCTLPPPSKHAPPGTHLW
                     EFTRDLLIPPELNFGLMKWENRHEGVPKFLRSEAVAyLwCQKKKNSNMTYEKLSRAMR
                     YYYKKRELLEKVIwGKKLVYKFGKNSSGWKEEEVIQSKN"
BASE COUNT      414 a    565 c    548 g    388 t
ORIGIN


alignment scores:
            Quality:  255.00            Length:    445
              Ratio:  1.471               Gaps:     17
 Percent Similarity:  42.759   Percent Identity:  22.989

alignment block:
 US-09-126-945B-2 x HSU73843   ..

 Align seq 1/1  to: HSU73843  from: 1  to: 1915

      25 ArqThrGlyLeuIleLysAlaAlaAlaIleGlyAlaValGlyLeuGluArqAr 41
         |||:::|||     :::|||||||||||      ::::::          ||
       3 AGAGCCGGGTAGGGAGCGCAGCGGCCAGATACCTCAGCGCTACCTGGCG 52

      41 qAspTrpSerProSerProPro.............................. 48
         |:::|||          ||||||
      53 GAACTGGATTTCTCTCCCGCCTGCCGGCCTGCCTGCCACAGCCGGACTCC 102

      49 .......................AlaThrProGluGlnGlyLeuSerAla 67
                               ||||||   |||      :::|||
     103 GCCACTCCGGTAGCCTCATGCCTGCAACCTGTGAG......ATTAGCAAC 146

      68 PheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerTrpAl 74
         ::::::   :::||||||||::   :::|||   ::::::::|||
     147 ATTTTAGCAACTACTTCAGTGCGATGTACAGCTCGGAGGACTCCACCCT 196

      74 aAlaLysAlaProGlyAlaSerSerArqGluGluProProGluGluProG 91
         |||:::::::||    |||
     197 GGCCTCTGTTCCCCCTGCT............................... 215

      91 luGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeu 107
                                       |||      ||::::   ||||||
     216 .........................GCCAGCCTTTGGGGCCGATGACTTG 239

     108 ValProGlyGlyLeuThrLeuGlu......GluHisSerLeuGluGlnVa 122
         |||       ||||||||||:::      :::   ||||||||   ::
     240 GTA.........CTGACCCTGAGCAACCCCCAGATGTCATTGGAGGGTAC 280

     122 lGln...SerMetValValGlyGluValLeuLysAspIleGluThrAlaC 138
         ::::   :::      :::||||||
     281 AGAGAAGGCCAGCTGGTTGGGGGAA........................... 305

     138 ysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsn 154
                                 ::::|||:::   ||||||   ::::::
     306 .......................CAGCCCCAGTTCCTGCTCCAACAGCCAG 332

     155 ValGlnLysTrpLeuLeuTrp......ThrGluHisGlnTyrArqLeuPr 169
         |||      ||:::   :::         ::::::::|||
     333 GTCTGGACTGGATCAGCTACCAAGTGGAGAAGAACAAGTACGACCAAG 382

     169 oProMetGlyLysAlaPheGluGluLeuAlaGlyLysGluLeuCysAlaM 186
         :::   :::      ::::::   |||      ||||||
     383 GGCCATGACTTCTCACGATGTGACATGGATGGCGCCACCCTCTGCAATT 432

     186 etSerGluGluGlnPheArq...GlnArqSerProLeuGlyGlyAspVal 201
         :::   |||:::::::|||       :::|||||   ||||||
     433 GTGCCCTTGAGGAGCTGCGTCTGGTCTTTGGGCCTCTG...GGGGACCAA 479

     202 LeuHisAlaHisLeu...AspIleTrpLysSerAla.....Al 213
         ||||||||||::|||   |||:::   :::|||:::          ::
     480 CTGCATGCCCAGCTGCGAGACCTCACTTCCAGCTTTCTGATGAGCTCAG 529

     213 aTrpMet.....................................LysGluArqT 219
         :|||:::                                   :::|||
     530 TTGGATCATTGAGCTGCTGGAGAAGGATGGCATGGCCTTCTCAGAGGCCC 579

     219 hrSerPr:::lyAlaIleHisTyrCysAlaSerThrSerGluGluSerTrp 235
         :::||||||   :::      :::      :::::|||
     580 TACACCCAGGGCCCTTGACCAGGGCAGCCCCTTGCCCAGGAGCTGCTG 629

     236 IhrAspSerGluValAspSer..............SerCysSer..... 245
         |||::::::      |||                |||||:::
     630 GACGACGGTCAGCAAGCCAGCCCCTACCACCCCGGCAGCTGTGGCAGG 679

     245 ..................................................... 245

     680 AGCCCCCTCCCCTGGCAGCTCTGACGTCTCCACCGCAGGGACTGGTGCTT 729

     245 ..................................................... 245

     730 CTCGGAGCTCCCACTCCTCAGACTCCGGTGGAAGTGACGTGGACCTGGAT 779

     245 ..................................................... 245

     780 CCCACTGATGGCAAGCTGTTCCCCAGCGATGGTTTTCGTGACTGCAAGAA 829

     246 ..............................GlyGlnPro.......... 248
                                      |||:::|||
     830 GGGGGATCCAAAGCACGGGAAGCGGAAACGAGGCCGGCCCCGAAAGCTGA 879

     248 ..................................................... 248

     880 GCAAAGAGTACTGGGACTGTCTCGAGGGCAAGAAGAGCAAGCACGCGCCC 929
```

```
     249 ......IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHi 264
                 ||||||||||:::|||||||||||||||||:::::   |||::
     930 AGAGGCACCCACCTGTGGGAGTTCATCCGGGACATCCTCATCCACCCGGA 979

     265 sSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheL 280
         :          ::::::::::|||   |||::::::::::|||:::||||
     980 GCTCAACGACGGCCTCATGAAGTGGGAGAATGGGCATGAAGGCGICTTCA 1029

     281 ysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsn 296
         ||:::      |||      |||||||:::|||||||||   :::|||:::
    1030 AGTTCCTGCGCTCCGAGGCTGTGGCCCAACTATGGGGCCAAAAGAAAAAG 1079

     297 ArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTy 313
         :::      |||:::|||:::||||||||||||||:::|||   |||||
    1080 AACAGCAACATGACCTACGAGAAGCTGAGCCGGGCCATGAGGTACTACTA 1129

     314 rLysLysGlyIleIleArgLysProAspIleSerGlnArgLeuValTyrG 330
         :|||:::   |||:::::::::   |||   :::::::|||||||||::
    1130 CAAACGGGAGATCCTGGAACGGGTGGAT...GGCCGGCGACTCGTCTACA 1176

     330 lnPhe 331
         ::|||
    1177 AGTIT 1181

seq_name: qb_pri0:HSU97156

seq_documentation_block:
LOCUS       HSU97156     1942 bp    mRNA            PRI       05-DEC-1997
DEFINITION  Homo sapiens epithelial specific ets protein (jen) mRNA, complete
            cds.
ACCESSION   U97156
VERSION     U97156.1  GI:2459796
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1942)
  AUTHORS   Andreoli,J.M., Jang,S.I., Chung,E., Cotlechia,C.M., Steinert,P.M.
            and Markova,N.G.
  TITLE     The expression of a novel, epithelium-specific ets transcription
            factor is restricted to the most differentiated layers in the
            epidermis
  JOURNAL   Nucleic Acids Res. 25 (21), 4287-4295 (1997)
  MEDLINE   97478557
REFERENCE   2  (bases 1 to 1942)
  AUTHORS   Markova,N.G., Andreoli,J.M., Jang,S.-I., Cotlechia,C.M. and
            Steinert,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-1997) Laboratory of CBDB, PI 1 of, NIAMS, NIH,
            Building 6, Room 425, Bethesda, MD 20892-2752, USA
FEATURES             Location/Qualifiers
     source          1. .1942
                     /organism="Homo sapiens"
                     /db_xref "taxon:9606"
     gene            1. .1942
                     /gene="jen"
     CDS             125. .1240
                     /gene "jen"
                     /codon_start=1
                     /product "epithelial-specific ets protein."
                     /protein_id "AAC51884.1"
                     /db_xref "GI:2459797"
                     /translation="MAATCEISNIFSNYFSAMYSSEDSTLASVPEAAIPSAGDLVETL
                     SNPQMSLESIEKASWTGEQRGWSKTQVTLWFSYQVEANSTDASAFTQSPCTMRSAIL
                     CNCALEELRLVFGPLGDQLHAQLRDLISSSSSDELSWTTELLEAIGMAEQLALDPCPPE
                     QGSPFAQELLDDGQQASPYHPGSCQAGAPSPGSSDVSIACTCASKSSHSSDSGCSDVD
                     LDPTDGKLFPSDGFRDCPKAGDVEHGKPEPCPPKELSETYWCTLCKSFNHAFPSCHLW
                     EFIRDILIHPELNEGLMKWENFHPCVPKFLPSEAVAQLWQEEENSNMIYFKLSKAMR
                     YYYKPEILEKVIGPKLVYRFGKNSSGWKEEFVLQSKN"
BASE COUNT      434 a     569 c     547 g     391 t       1 others
ORIGIN
```

```
alignment_scores:
           Quality:  255.00              Length:    435
             Ratio:  1.371                 Gaps:     17
 Percent Similarity:  42.759   Percent Identity:  22.989

alignment_block:
 US-09-126-945B-2 x HSU97156   ..

 Align seq 1/1  to: HSU97156  from: 1  to: 1942

      25 ArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgAr 41
         |||:::|||      :::||||||||||      ::::::          ||
       8 AGAGCCGGGTAGGGGAGCGGAGCGGGCAGATACCTCAGCGCTACCTGGCG 57

      41 qAspTrpSerProSerProPro..................................... 48
         |:::|||          ||||||
      58 GAACTGGATTTCTCTCCGCCTGCCGGCCTGCCTGCCACAGCCGGACTCC 107

      49 ..........................AlaThrProGluGlnGlyLeuSerAla 57
                                   ||||||   |||      :::|||
     108 GCCACTCCGGTAGCCTCATGGCTGCAACCTGTGAG......ATTAGCAAC 151

      58 PheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrpAl 74
         ::::::   :::||||||||:   :::|||   ||||||||||:|||
     152 ATTTTTAGCAACTACTTCAGTGCGATGTACAGCTCGGAGGACTCCACCCT 201

      74 aAlaLysAlaProGlyAlaSerSerArgGluGluProProGluGluProG 91
         |||:::::::||   |||
     202 GGCCTCTGTTCCCCCTGCT..................................... 220

      91 luGlnCysProValIleAspSerGluAlaProAlaGlySerLeuAspLeu 107
                                     |||      ||:::   ||||||
     221 ........................GCCACCTTTGGGGCCGATGACTTG 244

     108 ValProGlyGlyLeuThrLeuGlu......GluHisSerLeuGluInVa 122
         |||          ||||||||||:::     :::   ||||||||   ::
     245 GTA.........CTGACCCTGAGCAACCCCCAGATGTCATTGGAGGGTAC 285

     122 lGln...SerMetValValGlyGluValLeuLysAspIleGluThrAlaC 138
         ::::   :::      :::|||||
     286 AGAGAAGGCCAGCTGGTTGGGGGAA.......................... 310

     138 ysLysLeuLeuAsnIlethrAlaAspProMetAspTrpSerProSerAsn 154
                                 :::|||:::   ||||||   ::::::
     311 ......................CAGCCCCAGTTCTGGTCGAAGACGCAG 337

     155 ValGlnLysTrpLeuLeuTrp......ThrGluHisGlnTyrArgLeuPr 169
         |||      |||:::   :::           ::::::::::|||
     338 GTTCTGGACTGGATCAGCTACCAACTGGAGAAGAACAAGTACGACGCAAG 387

     169 oProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaM 186
         :::      :::      ::::::   |||      ||||||
     388 CGCCATTGACTTCTCACGATGTGACATGGATGGCGCCACCCTCTGCAATT 437

     186 etSerGluGluGluPheArg...GlnArgSerProLeuGlyGlyAspVal 201
           :::   |||:::::::|||         :::|||||   ||||||
     438 GTGGCCTTGAGGAGCTGCGTCTGGTCTTGGGCCTCTG...GGGGACCAA 484

     202 LeuHisAlaHisLeu...AspIleTrpLysSerAla.............Al 213
         ||||||||.::::||   |||:::   :::|||:::            ::
     485 CTCCATGCCCAGCTGCGAGACCTCATTCCAGCTCTTCTGATGAGCTCAG 534

     213 aTrpMet.....................................LysGluArgI 219
         :|||:::                                   :::|||
     535 TTGGATCATTGACGCTGCTGCAGAAGGATGGCATGGCTTCCCAGGAGCCC 584

     220 hr*****yAl*****erTyrC**ASerThrSerGluGlnSerTrp 245
         :::|||||||   :::      :::      ::::::|||
     585 TAGACCCAGGGCCCTTTGACCAGGGCAGCCCCTTTGCCCAGGAGCTGCTG 634

     236 ThrAspSerGluValAspSer..............SerCysSer..... 245
           |||:::::::      |||              ||||||:::
```

```
      645 GACGACGGTCAGCAAGCCAGCCCCTACCACCCCGGCAGCTGTGGCGCAGG 694

      245 ..................................................  245

      685 AGCCCCCTCCCCCGGCAGCTCTGACGTCTCCACCGCAGGGACTGGTGCTT 734

      245 ..................................................  245

      735 CTCCGAGCTGCCACTCCTCAGACTCCGTGGAAGTGACGTGGACCTGGAT 784

      245 ..................................................  245

      785 CCCACTGATGGCAAGCTCTTCCCCAGCGATGCTTTCGTGACTGCAAGAA 834

      246 ..............................GlyGluPro..........  248
                                        |||:::|||
      835 GGGGGATCCCAAGCACGGGAAGCGGAAACGACGCCGCCCCCGAAAGCTGA 884

      248 ..................................................  248

      885 GCAAAGACTACTGGGACTGTCTCGAGGGCAAGAAGACCAACGACGCCCC 934

      249 ......IleHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHi 263
                |||||||||||:|||:||||||||||||||:::   ||||:
      935 AGAGGCACCCACCTGTGGGAGTTCATCGGGACATCCTCATCCACCCGGA 984

      264 sSerTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePhe 280
          :          ||||||||||||   |||||||||||||||||||||||
      985 GCTCAACGAGGGCCTCATGAAGTGGGAGAATCGGCATGAAGGCGTCTTCA 1034

      280 ysIleGluAspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsn 296
          ||||:    |||     |||||||:||||||||    :::||||::
     1035 AGTTCCTGCGCTCCGAGGCTGTGGCCCAACTATGGGGCCAAAACAAAAAG 1084

      297 ArgProAlaMetAsnTyrAspLysLeuSerArgSerIleArgGluTyrTy 313
          :::     |||:::|||:::|||||||||||||||:::::|||   |||||
     1085 AACAGCAACATGACCTACGAGAAGCTGAGCCGGCCATGAGCTACTACTA 1134

      314 rLysLysGlyIleIleArgLysProAspIleSerGluArgLeuValTyrG 330
          ||||:::   |||:::::::::   |||   :::::|||||||||||||:
     1135 CAAACGGGAGATCCTGCAACGGCTGGAT...GGCCGGCGACTCGTCTACA 1181

      330 lnPhe 331
          ::|||
     1182 AGTTT 1186

seq name: qb_pri:BC003569

seq documentation_block:
LOCUS       BC003569     1983 bp    mRNA            PRI       16-MAR-2001
DEFINITION  Homo sapiens, E74 like factor 3 (ets domain transcription factor,
            epithelial-specific ), clone MGC:2088, mRNA, complete cds.
ACCESSION   BC003569
VERSION     BC003569.1  GI:13097734
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1983)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH MGC Project Url: http://mgc.nci.nih.gov
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: NCI/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ron Grin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Mitanada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at  http://image.llnl.gov
            Series: IRAL Plate: 8 Row: j Column: 17.
FEATURES             Location/Qualifiers
     source          1..1983
                     /organism="Homo sapiens"
                     /db_xref="LocusID:1999"
                     /db_xref="taxon:9606"
                     /clone="MGC:2088 IMAGE:3537846"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B_R"
                     /note="Vector: pOTB7"
     CDS             122..1237
                     /codon_start=1
                     /product="E74 like factor 3 (ets domain transcription
                     factor, epithelial-specific )"
                     /protein_id="AAH03569.1"
                     /db_xref="GI:13097735"
                     /translation="MAATCEISNIFSNYFSAMYSSEDSILASVPFAAIFGADDLVELL
                     SNPQMSIFGIEKASWLGEQPQFWSKTQVELQWISYQVEKNKYDASAIDFSRCDMDGAIL
                     CNCALEELKLVFGPLGDQLHAQLRDLTSSSSDELSWIIELLEKDGMAFQEALDFGPFD
                     QGSPFAQELLDDGQQASPYHPGSCGAGAPSFGSSDVSTAGTGASKSSHSSDSCGSDVD
                     LDPIDGKLFPSLGFRICKKGDPKHGKRKKKGKPPKKLSKEYWE* LEGKASKHAIFGIHLW
                     EFIKDILIHPELNEGLMKWENRHEGVFKFLRSEAVAQLWGQKKKNSNMIYEKLSRAMR
                     YYYAKEILEKVDGKKLVYKFGKNSSGWKEEEVLQSKN"
BASE COUNT       480 a    566 c    548 g    389 t
ORIGIN


alignment scores:
            Quality:  255.00              Length:    435
              Ratio:  1.371                 Gaps:     17
 Percent Similarity:  42.759   Percent Identity: 22.989

alignment_block:
 US-09-126-945B-2 x BC003569  ..

 Align seq 1/1  to: BC003569  from: 1  to: 1983

      25 ArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgAr 41
         |||:::|||      :::|||||||||      :::::          ||
       5 AGAGCCGGGTAGGGGAGCGCAGCGGCCAGATACCTCAGCGCTACCTGGCG 54

      41 qAspTrpSerProSerProPro..............................  48
         |:::|||          ||||||
      55 GAACTGGATTTCTCTCCCGCCTGCCGGCCTGCCTGCCACAGCCGGACTCC 104

      49 .......................AlaThrProGluGlnGlyLeuSerAla  57
                                 ||||||   |||      :::|||
     105 GCCACTCCCGCTAGCCTCATGGCTGCAACCTGTGAG......ATTAGCAAC 148

      58 PhetyrLeuSertyrPheAspMetLeuTyrProGluAspSerSerTrpAl  74
         ::::::   :::||||||||:   :::|||   ::::::::|||
     149 ATTTTACACAACTACTCAGTGCGAIGTACAGCTCGGAGGACTCCACCCT 198

      74 aAlaLysAlaProGlyAlaSerSerArgGluGluProProGlnGluProG  91
         |||:::::|||   |||
     199 GGCCTCTGTTCCCCCTGCT.............................. 217

      91 lnGlnCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeu 107
```

```
                                  |||       ||||::   |||||
218 ..........................GCCACCTTTGGGGCCGATGACTTG 241

108 ValProGlyGlyLeuThrLeuGlu......GluHisSerLeuGluGlnVa 132
    |||          ||||||||||::      :::   ||||||||   ::
242 GTA.........CTGACCCTGAGCAACCCCCAGATGTCATTGGAGGGTAC 282

122 lGln...SerMetValValGlyGluValLeuLysAspIleGluThrAlaC 146
    ::::   :::       :::|||||
283 AGACAAGGCCAGCTGGTTGGGGGAA........................... 307

138 ysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsn 154
                        ' :::|||:::   ||||||   ::::::
308 ......................CAGCCCCAGTTCTGGTCGAAGACGCAG 334

155 ValGlnLysTrpLeuLeuTrp......ThrGluHisGlnTyrArgLeuPr 169
    |||      ||:::   :::          ::::::::|||
335 GTTCTGGACTGGATCAGCTACCAACGTGGAGAAGAACAAGTACGACGCAAG 384

169 oProMetGlyLysAlaPheGlnGluLeuAlaGlyLysGluLeuCysAlaM 186
      :::      :::      ::::::   |||      ||||||
385 CGCCATTGACTTCTCACGATGTCACATGGATGGCGCCACCCTCTGCAATT 434

186 etSerGluGluGlnPheArq...GluArqSerProLeuGlyGlyAspVal 201
    :::   |:|:::::|||         :::||||||   ||||||
435 GTGCCCTTGAGGAGCTGCGTCTGGTCTTTGGCCTCTG...GGGGACCAA 481

202 LeuHisAlaHisLeu...AspIleTrpLysSerAla.............Al 218
    |||||||||:::|||   ||:::   :::|||:::          ::
482 CTCCATGCCCAGCTGCGAGACCTCACTTCCAGCTCTTCTGATGAGCTCAG 531

218 aTrpMet................................LysGluArqI 229
    ::::::                               ::::||
532 TTGGATCATTGAGCTGCTGGAGAAGGATGGCATGGCCTTCCAGGAGGCC 581

229 hrSerProGlyAlaIleHisTyrCysAlaSerThrSerGluGluSerIrp 245
      :::||||||   :::        :::       :::::|||
582 TAGACCCAGGGCCCTTTGACCAGGGCAGCCCCTTTGCCCAGGAGCTGCTG 631

236 ThrAspSerGluValAspSer                  SerCysSer      247
      ||:::::::      |||                 ||||||||:
632 GACGACGGTCAGCAAGCCAGCCCCTACCACCCCGGCAGCTGTGGCGCAGG 681

245 ..................................................... 245

682 AGCCCCTCCCCTGCCAGCTCTGACGTCTCCACCCCACCGACTTCGTCTTT 731

245 ..................................................... 245

732 CTCGGAGCTCCCACTCCTCAGACTCCGCTGCAACTGACGTCGACCCTGCAT 781

245 ..................................................... 245

782 CCCACTGATGGCAAGCACTTCCCCAGCGATGGTTTTCGTGACTGCAAGAA 831

246 ............................GlyGlnPro.......... 248
                              ||||:|||
832 GGGGGATCCCAAGCACGGGAAGCCGAAACGAGGCCGGCCCCGAAAGCTGA 881

248 ..................................................... 248

882 GCAAAGAGTACTGGGACTGTCTCGAGGGCAAGAAGAGCAAGCACGCGCCC 931

249 .....IleHisLeuTrpGlnPheLeuLysGluLeuLeuLysPro:Hi 263
          |||||||||:::|||:::::::::::::::::::::   ||||:
932 AGACGCACCCACCTGTGGGAGTTCATCCGGGACATCCTCATCCACCCGA 981

263 sSerTyrGlyArqPheIleArqTrpLeuAsnLysGluLysGlyIlePheL 280
    :   :::::::::|||   |||:::::::::::|||:::|||||
982 GCTCAACGAGGGCCTCATGAAGTGGGAGAATCGGCATGAAGGCGTCTTCA 1031

280 yeIle:lgAspSerAlaGlnValAlaArqLeuTrpGlyIleArqLyeAsn 299
    |||::      |||      |||||||:::|||||||||   :::|||:::
```

```
     1032 AGTTCCTGCGGCTCCGAGGCTGTGGCCAACTATGGGGCCAAAAGAAAAAG 1081

      297 ArqProAlaMetAsnTyrAspLysLeuSerArqSerIleArqGlnTyrLy 313
          :::      |||:::|||:::|||||||||||||:::::::|||   ||||
     1082 AACAGCAACATGACCTACGAGAAGCTGAGCCGGCCCATCAGGTACTACTA 1131

      313 rLysLysGlyIleIleArqLysProAspIleSerGlnArqLeuValTyrG 330
          ||||:::   |||:::::::::   |||   :::::|||||||||||:
     1132 CAAACGGGAGATCCTGGAACGGGTGGAT...GGCCGGCGACTCGTCTACA 1178

      330 lnPhe 331
          ::|||
     1179 AGTTT 1183

seq_name: qb_ov:CHKETS1

seq_documentation_block:
LOCUS       CHKETS1       2764 bp    mRNA            VRT       28-SEP-1989
DEFINITION  Chicken protein p54 (ets-1) mRNA, complete cds.
ACCESSION   M22462 M23688
VERSION     M22462.1  GI:211752
KEYWORDS    p54 protein.
SOURCE      G.gallus (2 week) thymus, cDNA to mRNA, clone lambda C111.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 2764)
  AUTHORS   Watson,D.K., McWilliams,M.J. and Papas,T.S.
  TITLE     A unique amino-terminal sequence predicted for the chicken
            proto ets protein
  JOURNAL   Virology 167, 1-7 (1988)
  MEDLINE   89045631
COMMENT     Draft entry and computer readable sequence for [Virology 167, 1-7
            (1989)] kindly provided
            by J.A.Lautenberger 21-FEB-1989.
FEATURES             Location/qualifiers
     source          1. .2764
                     /organism "Gallus gallus"
                     /db_xref "taxon:9031"
     CDS             1449. .2764
                     /note "p54 protein"
                     /codon_start=1
                     /protein_id "AAA48764.1"
                     /db_xref "GI:211753"
                     /translation="MKAAVDLEAFLLIKEIEEKVSTLQEFSTLMEQAQVQLLVSSEEM
                     MSQALKALFSCFAKFQQPICTDKDPQQWIFTHVPEWVMWAVNEFSLFQVIDLQEQMNG
                     AALCALGKEQFTLELAPDFVGDTLWEHLETLQKEEAKPYPANGVNAAYPESKYISQYFI
                     SYGIEHAQCVPPSEFSEPSFTTESYQTLHPISSEELLSLKYFNEDYPSVILRDPVQTDS
                     LQTDYFTLTQFVVTPDNMCMCPASPGKIAGQDSFPSIFSYDSCDRLTQSWSSQSSFQS
                     LQRVPSYDSFDSPDYPAALPNHLPEGIFEDYVVDPALMNEDGPVTFAAALAGYDGSGP
                     TQLWQFLLELLTDKSCQSFISWIGDHWEKLSDFDFEVARPWGKEKKNFEKMNYFKLSEG
                     LRYYYDRNTLTHKTAGSPYVYPPVCTDQQSTLGYTPEELHAMLDVKFDADE"
BASE COUNT      696 a    760 c    876 g    432 t
ORIGIN


alignment_scores:
            Quality:  254.50              Length:    400
              Ratio:  1.611                 Gaps:     15
 Percent Similarity:  39.500   Percent Identity: 24.000

alignment_block:
 US-09-126-945B-2 x CHKETS1   ..

 Align seq 1/1  to: CHKETS1  from: 1  to: 2764

      105 LeuAspLeuValPro.................GlyGlyLeuThrLeuGl 115
          :::||||||   |||                :::   :::   |||
     1499 AlaGlaCiCTTCCCGiCCCCGaIAtGGAAIGtGCAGAtGtGCCtTtGtt 1548

      115 sGlnHisSerLeuGluGlnValGluSerMetValValGlyGlnValLeuL 142
                  |||   ::::::::::::|||:::::::::     :::::::
```

```
1549 AACCCCCAGCAGCAAGGAAATGATGTCTCAGGCACTGAAAGCCACCTTCA 1598

 132 ysAspIleGluThrAlaCysLysLeuLeuAsnIleThrAlaAspProMet  148
     ::   :::            :::   |||:::|||      |||||||:::
1599 GTGGCTTCGCAAAGGAGCAGCAGCGGCTGGGAATCCCCAAAGATCCCCAG 1648

 149 AspTrpSerProSerAsnValGluLysTrpLeuLeuPheThrGluHisGl  165
     :::|||:::   :::::|||:::   |||:::::|||:::   :::::
1649 CAGTGGACAGAGACGCACGTGCGGACTGGCTGATGTGGGCAGTGAATGA 1698

 165 nTyrArgLeuProProMetGlyLysAlaPheGluGlu......LeuAlaG  180
     ::::   |||         |||       ||||||:::       :::   |
1699 GTTCAGCCTG......AAGGGACTGGATTTCCAGAAGTTCTGCATGAACG 1742

 180 lyLysGluLeuCysAlaMetSerGluGluGluPheArgGluArgSerPro  196
     ||     ||||||||||:::::::::|||   |||   :::   :::|||
1743 GAGCTGCCCTGTGCGCCCTGGGCAAGGAGTGCTTCCTGGAGCTAGCGCCT 1792

 197 ...LeuGlyGlyAspValLeuHisAlaHisLeuAspIleTrpLys..... 210
        :::   ||||||:::|||      ||||||:::|||   :::
1793 CACTTTGCGGACATATCCTTTGCGAACACCTGGAGATCTTGCAGAAAGA 1842

 211 ..............................SerAlaAlaTrpMetLysG  217
                                   :::||||||:::   ::::
1843 AGAGGCAAAACCATACCCAGCAAATGGAGTGAATGCAGCGTATCCAGAAT 1892

 217 lyArg...ThrSerProGlyAlaIleHisTyr................. 226
     ::|||   ||||||      |||   |||
1893 CCCGCTATACTTCAGACTACTTCATTAGTTATGGCATCGAGCACGCACAG 1942

 227 CysAlaSerThrSerGlu.....................GluSerIr  235
     |||:::      ||||||                     |||||||::
1943 TGCGTGCCTCCCTCCGAGTTCTCTGAGCCCAGCTTCATCACAGAGTCCTA 1992

 235 p.................................................. 235
     :
1993 CCAGACCCTCCATCCCATCAGCTCGGAAGAGCTTCTGTCCCTCAAGTACG 2042

 236 ..........................................ThrAspSer  238
                                               |||||||||
2043 AGAACGACTATCCCTCAGTCATCCTTCGTGACCCCGTCCAGAGGACTCC 2092

 239 ...............................GluVal.........AspSe  242
                                   ||||||         ||||:
2093 CTGCAGACAGACTACTTCACAATCAAGCAAGAAGTGGTAACGCCAGACAA 2142

 242 rSerCysSerGlyGln................................. 247
     :  |||  |||:::
2143 CATGTGCATGGGACGTGCCAGTCGAGGTAAACTGGGTGGCCAGGACTCCT 2192

 247 ................................................. 247

2193 TTGAGAGCATAGAGAGCTACGACAGCTGTGACGGCCTGACACAGTCCTGG 2242

 247 ................................................. 247

2243 AGCAGCCAGTCCTCCTTCCAGAGCCTGCAGCGGCGTCCCCTCCTACGATAG 2292

 247 ................................................. 247

2293 CTTTGACTCAGAGGACTACCCCGCCGCCCTGCCCAACCACAAGCCCAAGG 2342

 247 ................................................. 247

2343 GCACCTTCAAGGACTATGTTCGACATCGGCTGACATGAACAAGGACAAG 2392

 248 ..............................................ProIl 249
                                                   |||||
2393 CCTGTCATTCCTGCCGCTGCCCTCGCCGGCTACACAGCAGTGGACCCAT 2442

 249 eHisLeuTrpGlnPheLeuLysGluLeuLeuLeuLysProHisSerTyrG 266
     |:::|||||||||||||||   |||   ||||||          |||
2443 CCAACTGTGGCAATTCCTGCTGGAG...CTGCTCACTGACAAGTCCTGTC 2489

 266 lyArgPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGlu 282
         ||||||   |||   :::   :::          ||||||||:::
2490 AGTCCTCACAGCTGGACGGGT...GATGGCTGGGACTTCAAGCTTCC 2536

 283 AspSerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAl 299
     |||       :::||||||||||   ||||||   ||||||||||:::|||
2537 GATCCAGATCAGCTGCCCAGGCGGTGGGGCAAGAGGAAAAACAAGCCCAA 2586

 299 aMetAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysG 316
      ||||||||||:::|||||||||||||||:::::|||   ||||||   |||:
2587 GATGAACTATGAGAAGCTGAGCCGTGGTCTGCGTTACTATTACGACAAGA 2636

 316 lyIleIleArgLysProAspIleSerGluArgLeuValTyrGlnPheVal 332
     ::||||||:::|||          ::::::|||   ||||||:::||||||
2637 ACATCATCCACAAG...ACGGCCGGCAAGCGCTACGTCTACGCTTCGTC 2683

seq name: qb in3:CELC24A1

seq documentation_block:
LOCUS       CELC24A1     39248 bp    DNA              INV       16 SEP 1997
DEFINITION  Caenorhabditis elegans cosmid C24A1.
ACCESSION   AF024491
VERSION     AF024491.1  GI:2394415
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 39248)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vandlin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 39248)
  AUTHORS   Connell,M.
  TITLE     The Sequence of C. elegans cosmid C24A1
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 39248)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11 SEP 1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University,
                  St. Louis, Mo  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 1RQ, England
                  e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  this sequence may not be the entire insert of this clone,
            it may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone
```

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is W02B3, 200 bp overlap; 5' cosmid is [illegible], 3997 bp overlap. Actual start of this cosmid is at base position 197 of CELC24A1; 3' cosmid end is at 4572 of CELF17H7

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder(P. Green and L. Hillier, ms in preparation).

```
FEATURES             Location/Qualifiers
     source          1..39248
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="C24A1"
     gene            complement(320..3920)
                     /gene="C24A1.3"
     CDS             complement(join(320..823,867..984,1234..1359,1402..1448,
                     1495..1695,1743..2224,2277..2627,2889..3121,3191..3629,
                     3793..3920))
                     /gene="C24A1.3"
                     /note="contains similarity to ankyrin repeats and protein
                     kinase motifs"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB70312.1"
                     /db_xref="GI:2394418"
                     /translation="MKIFSILLFxxIxIGYSVVExELxxVExExNxWVxVxIAAFE
                     KSLEDEFQHMCPENAFIKSTQIISIFHLIPAPSPSQPPKIxEFIENLPKxSSIIGS
                     SQSANGFTPLHIAIYRGDVAILKALIATKLVDLDQSGRHLLFALHLAAMIGDSEMLI
                     LLNSGANIHVTDFVHPTALHCATYFGQENAVRTLISASANLNIQDAVNRFPLHLAAAE
                     GLISIFELLLEAKADPLLADDEGNQALHYAAKSGSLVILNMLIKQVRGTNDRIGARNL
                     YGDTALHLSCYSGRLDIVKSILDSSPINIVNMLNVFSxTPLHAAVIGGxxELLVSxGM
                     KYPGVDPNYQGQDGHTALHSACYHGHLRIVQYLLENGADQSLASKAFEGGALRQQAGP
                     GTNRPSEVASAIMALSRPDLIExSNAPYNxTVLxLxQQLIxVxWAYExCHxMxVALxxx
                     AARTVEGDVCSEYSSGESSYTPLPSPMGRLTSLIRDKADLLQLRSALPAPFHLCLAEI
                     EFQESIGSGSPGKVYKGTYRGELVAVExxYxAMAIQxxKSExIxMLxxEVSLLSRLAHPNV
                     VAFVGTSLDDPSQFAIITEFVENGSLERRENGERKNYRVMIxFAFDIKRGLLVARGMRY
                     LHESAAKPVIHRDLNSRNILIHADMRAKVSVVADFGLSRFVQxxExxxENLLExQxxNLPWMAP
                     EVFSQSGKYQMKVDVFSFALVLWEIHTAKIPFSHLKPAAAAAFMIYKPQPPILPNQPI
                     AQFPAHILSLIFQAWRPESSSRFTFVEIVAILEFRVISIHIPIISAPSTVSxxDLSxWDx
                     LSVAPPPASKFPPILSALHGIAAIGIVEELRQRIDNNGYVINKS"
     gene            complement(8283..8801)
                     /gene="C24A1.4"
     CDS             complement(8283..8801)
                     /gene="C24A1.4"
                     /note="contains similarity to mariner transposase"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB70313.1"
                     /db_xref="GI:2394419"
                     /translation="MRTIKAFALxLLGIIIWKxIxxIKAIxAAGxAxNHIAYIAGHLxYVNH
                     TKEKxWVYVExLRAKTLKxxLHExxVIxxxxWWxxIGVIxxxxxxxxxxxxxxxSAxxxxxxx
                     EKVVHARQLHKPRGSKLLLLHDNARPHIAVKIRQKLQFSGVEVLPHPPYSPDLAPSDFFLFGYLKSKLRGTRFQTIDELKQ
                     KCCPLFCCPLHL"
     gene            14193..23291
                     /gene="C24A1.2"
     CDS             join(14193..14457,14917..15057,15857..15956,16818..17000,
                     18450..18507,18551..18620,21048..21179,21943..22307,
                     22977..23027,23181..23291)
                     /gene="C24A1.2"
                     /note="contains similarity to FTS domains; coded for by C.
                     elegans cDNA yk158e6.3, coded for by C. elegans cDNA
                     yk151a10.3, coded for by C. elegans cDNA yk158e6.5, coded
                     for by C. elegans cDNA yk151a10.5"
                     /codon_start=1
                     /protein_id="AAB70311.1"
                     /db_xref="GI:2394417"
                     /translation="MIQTQGQPRPFPLLLVLPIAISPLNISSSNISSKSSRTNNNNEN
                     FNDEEDDGRDLDVEMKTESEWPRTERPTPKYADGRRISAPAVLAIDGLPEIVINFMERP
                     DFCTSTIVGSEFHAEFGADEQSRPWCWHRLDTRNVSVKFNDTFSKICSALISTLSYS
                     TPAYRNPTISVAIVGLALLSACAECERCRMSTFQPQSGKIVTGLLFDRKTPNAEFA
                     QFFFGVPINSFGQNDAIQRWLQSALLCPTTSASRPNLLTTHQISVQPQPTAPQILIA
                     APESSIVEIPCGDLLICLPSNPSTLEHPYLLDTSPKFSEFNFAPLAQRTICEEGKL
                     GQVIYILWEFLLPLLQPMQYSPKFIRWIDQAKCIERLVDSKAVSFLWCMHPNSI GMNYE
                     TMCRALPYYYQPCTLQKVDQPLVYPFVPLPSDVASSVSQFIPVCKFFTSEVWE SGS
                     CESADSCDSPKISPISLKLMLQPATSASDEA"
     gene            32785..33481
                     /gene="C24A1.1"
     CDS             join(32785..33010,33602..33693,34211..34481)
                     /gene="C24A1.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB70310.1"
                     /db_xref="GI:2394416"
                     /translation="MGKSFFLIHFLNFSVVxIxxxxKMAVxVxNGxKFIAGSKIFSIL
                     KMLSSRTSSIILILAILVATMAVAQCRNIQYIVEEMIIPAAAFPYAQLWDIIPHKRVPSA
                     CLMMLIIxLNxxGxxNNxxNxSVLKWSNLxGCAIAVHKLELIGVVxxSRIIENFKNE
                     AN"
BASE COUNT     12568 a    7038 c    6868 g   12774 t
ORIGIN
```

```
alignment scores:
            Quality:  252.50              Length:      88
              Ratio:  4.556                 Gaps:        1
 Percent Similarity:  80.682   Percent Identity: 51.136

alignment block:
 03-09-126-945b-2.x CELC24A1    ..

 Align seq 3/3   top: CELC24A1  from: 1  to: 39248

     246 GlyGlnProIleHisLeuTrpGlnPheLeuLysGlnLeuLeuLeuLysPr 282
         ||||||      :::|||||:::|||||   :::|||||
   22116 GGACAACATCACCTACATTTTGCLAAATTTTGTTAACACCCCCCACAGGA 22165

     283 oHisserTyrGlyArgPheIleArgTrpLeuAsnLysGluLysGlyIleP 329
         :::     :::|||||||:::||||||:::  |||||||||
   22166 CCAATAATCCGCCAAACTGTATTAAAATCCATATCATCAACCAAAACCAATTT 22215

     330 heIyslleGlnAspSerAlaGlnValAlaArgGlnProIlyIleArLys 376
         ||||||:::   ||||||      |||:::|||||||||||||:::|||
   22216 TTAAATTGTTCCATTTGAAACTAATTATTCTTGTTGTCAAATCCATAAG 22265

     376 AsnArgProAlaMetAsnPyrAspLysLeuSerArgSerIleArgGlnIy 412
         |||:::|||:::||||||||||:::  :::::|||:::::|||   ||
   22266 AATAACCCTGCAATCAATTATGACATAATCGCAACACCCCTCAGATATTA 22315

     412 rTyrIysIysGlyIleIeArgLysProAspIleSerGlnArgLeuValI 329
         ||||:::::||||||:::::|||   |||   :::||||||||||||
   22316 ::AT'AAA'.'.'.'A:::'::. A'AAA:::.'A:::.:.: AAA'A:.:'''' 22364

     329 yrGlnPheValHis 333
         |||:::|||||||||
   22365 ACACATCTGCCAT 22376

seq_name: qb_pat2:196206

seq_documentation block:
LOCUS       196206          1604 bp    DNA             PAT       01-DEC-1998
DEFINITION  Sequence 43 from patent US 5734039.
ACCESSION   196206
VERSION     196206.1  GI:3940676
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1604)
  AUTHORS   Calabretta,B. and Ck::ki,I.
  TITLE     Antisense oligonucleotides targeting cooperating oncogenes
```

```
  JOURNAL   Patent: US 5744039-A 44 31 MAR 1998;
FEATURES             Location/Qualifiers
     source          1..1604
                     /organism "unknown"
BASE COUNT       418 a     443 c     408 g     335 t
ORIGIN


alignment scores:
            Quality:  251.50              Length:    540
              Ratio:  1.251                 Gaps:     22
 Percent Similarity:  37.925   Percent Identity: 21.698

alignment block:
 US 09 126 945B 2 x 196206   ..

 Align seq 1/1  to: 196206  from: 1  to: 1604

      24 SerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluAr 40
         :::|||:::|||   |||:::   |||   |||:::
      10 GCGAGGGCCGGGCAGGAGGAGCGGGCCCGCCGCCGGG.............. 45

      40 qArqAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerA 57
          |||   |||:::||||||       :::::::|||||||||||
      46 ,CGAGGCTGGGACCCGAGCGGCGCTCACTTCGCCCGCAAAGTCCC....... 87

      57 LaPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrp 73
                                              :::   |||
      88 ..................................AACTTCCCCTGC 99

      74 AlaAla...LysAlaProGly..AlaSerSerArgGluGluPro..... 86
         :::|||   :::||||||||:::  |||   :::::||||:::
     100 AGTGCCGGGCGCGCACCGTCCGGGCGCGGGGCAAAGAAAGGCAGCGGCAA 149

      87 ..............................ProGluGluProGluGluCy 98
                                        |||   :::|||:::   ||
     150 TTTGAGATTTTTGGGAAGAAAGTCGGATTCCCCCGTCCCCTTCCCCCTG 199

      99 s.....................................................P 94
         |                                                    |
     200 TTACTAATCCTCATTAAAAAGAAAAACAACAATAACTGCAAACTTGCTAC 249

      94 roValIleAspSerGlnAlaProAla.........GlySerLeuAspLeu 107
         ||        |||   :::|||:::         :::::::::||||||
     250 CATCCCGTACGTCCCCCACTCCTGGCACCATGAAGGCGGCCGTCGATCTC 299

     108 ValPro.................................................. 109
           |||
     300 AAGCCGACTCCTCACCATCATCAAGACGGAAAAAGTCGATCTGGAGCTTTT 349

     110 ......................GlyGlyLeuThrLeuGluGluHisSerL 119
                                 :::   :::   |||         |||
     350 CCCCTCCCCCGGATATGGAATGTGCAGATGTCCCACTATTAACTCCAACCA 399

     119 euGluGlnValGlnSerMetValValGlyGluValLeuLysAspIleGlu 135
         :::::::::::::|||:::::::::     :::::::::  :::
     400 GCAAAGAAATGATGTCTCAAGCATTAAAAGCTACTTTCAGTGGTTTCACT 449

     136 PhrAlaCysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerPr 152
              :::   |||:::|||      ||||||   :::|||:::
     450 AAAGAACAGCAACGACTGGGCATCCCAAAAGACCCCCGGCAGTGGACAGA 499

     152 oSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGluTyrArgLeuP 169
         :::::|||:::   |||:::::|||:::   :::::::::   |||
     500 AACCCATGTTCGGGACTGGGTGATGTGGGCTGTGAATGAATTCAGCCTG. 548

     169 roProMetGlyLysAlaPheGluGlu......LeuAlaGlyLysGluLeu 183
                |||       |||||||:::       :::   |||       |||
     549 .....AAAGGTGTAGACTTCCAGAAGTTCTGTATGAATGGAGCAGCCCTC 594

     184 CysAlaMetSerGluGluGlnPheArgGlnArgSerPro...LeuGlyGl 199
         ||||||:::::::::::   |||   :::   :::|||   :::   ||
     594 TGCGCCCTGGGTAAAGACTGCTTTCTCGAGCTGGCCCCAGACTTTGTTGG 643

     199 yAspValLeuHisAlaHisLeuAspIleTrpLys..................... 210
         ||||:::|||       |||||||:::|||   :::
     644 GGACATCTTATGGGAACATCTAGAGATCCTGCAGAAAGAGCATGTCAAAC 693

     211 .....................SerAlaAlaTrpMetLysGluArq...Thr 219
                               :::   ||||::   ::::::|||   |||
     694 CATATCAACTTAAATGGAGTCAACCCAGCCTATCCAGAATCCCCTATACC 743

     220 SerProGlyAlaIleHisTyr.....................CysAlaSer15 230
         |||          |||   |||                    ||||::
     744 TCGGATTACTTCATTAGCTATGGTATTGAGCATGCCCAGTGTGTTCCACC 793

     230 rSerGlu.........................GluSerTrp.......... 235
          ||||||                       |||||||:::
     794 ATCGGAGTTCTCAGAGCCCAGCTTCATCACAGAGTCCTATCAGACGCTCC 843

     235 ..................................................... 235

     844 ATCCCATCAGCTCGGAAGAGCTCCTCTCCCTCAAGTATGAGAATGACTAC 893

     236 ..............................ThrAspSer............ 238
                                          |||||||:::
     894 CCCTCGGTCATTCTCCGAGACCCTCTCCAGACAGACACCTTGCAGAATGA 943

     239 ....................GluVal.........AspSerSerCysSerG 246
                             |||||          ||||::   |||   |
     944 CTACTTTGCTATCAAACAAGAAGTCGTCACCCAGACAACATGTGCATGG 993

     246 lyGln.................................................. 247
         ||:::
     994 GGAGGACCAGTCGTGGTAAACTCGGGGGCCAGGACTCTTTTGAAAGCATA 1043

     247 ..................................................... 247

    1044 GAGAGCTACGATAGTTGTGATCGCCTCACCCAGTCCTGGAGCAGCCAGTC 1093

     247 ..................................................... 247

    1094 ATCTTTCAACAGCCTGCAGCGTCTCCCTCCTATGACAGCTTCGACTCAG 1143

     247 ..................................................... 247

    1144 AGGACTATCCGGCTGCC rGCC AACC ACAAGCG AAGGG ACC TTCAAG 1193

     247 ..................................................... 247

    1194 GACTATGTGCGGGACCGTGCTGACCTCAATAAGGACAAGCCTGTCATTCC 1243

     248 ..................................ProTyrHisLeuTrpG 253
                                          |||||||:::|||||||
    1244 TGCTGCTGCCCTAGCTGGCTACACAGGCAGTGGACCAATCCAGCTATGGC 1293

     253 lnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 269
         |||||||   |||   ||||||          |||          ||||||
    1294 AGTTTCTTCTGGAA...TTACTCACTGATAAATCCTGTCAGTCTTTTATC 1340

     270 ArgTrpLeuAsnLysGlyLysGlyIleProLysIle:sAspSerAlaGl 286
         |||   :::   :::          |||||||:::::|||      ::
    1341 AGCTGGACAGGA...GATGGCTGGGAATTCAAACTTTCTGACCCAGATGA 1387

     286 nValAlaArqLeuTrpGlyIleArqLysAsnArqProAlaMetAsnTyrA 303
         :|||||||||   ||||||   |||||||||:::|||   |||||||||:
    1388 GGTGGCCAGGACATGGGCAAAGAGGAAAAACAAACCTAAGATGAATTATC 1437

     303 spLysLeuSerArqSer:IleArqGlnTyrTyrLysLysGlyIleIleArq 319
         ::||||||||||||:::::|||   ||||||   |||:::||||||:::
    1438 AGAAACTGAGCCGTGGCCTACGCTACTATTACGACAAAAACATCATCCAC 1487

     320 LysProAspIleSerGlnArqLeuValTyrGlnPheVal 352
         |||         :::::|||   ||||||:::||||||
    1488 AAG...ACAGCGGGGAAACGCTACGTGTACCGCTTTGTG 1523
```

```
seq_name: gb_pr9:HSCETS1

seq_documentation_block:
LOCUS       HSCETS1      1604 bp    mRNA            PRI       24-APR-1993
DEFINITION  Human DNA for c-ets-1 proto-oncogene.
ACCESSION   X14798
VERSION     X14798.1  GI:29881
KEYWORDS    alternative splicing; c-ets-1 gene; c-ets-1 proto-oncogene; ets
            oncogene; oncogene; proto-oncogene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1604)
  AUTHORS   Reddy,E.S. and Rao,V.N.
  TITLE     Structure, expression and alternative splicing of the human c-ets-1
            proto-oncogene
  JOURNAL   Oncogene Res. 3 (3), 239-246 (1988)
  MEDLINE   89084219
FEATURES             Location/Qualifiers
     source          1. .1604
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="lambda 4"
                     /cell_line="MOLT-4"
                     /clone_lib="lambda gt10 cDNA"
     CDS             279. .1604
                     /note="c-ets-1a protein (AA 1 441)"
                     /codon_start=1
                     /protein_id="CAA32904.1"
                     /db_xref="GI:29882"
                     /db_xref="SWISS-PROT:P14921"
                     /translation="MEAAVIHEPLQTFLLKEFVLSDPAVGNLIWVTDRHTGLRNKIVSLDGSKGM
                     MSQALKATESGFTKFQQWRTGIPKDPRQWTFTHVRTWVMWAVNEFSLKGVDFQKFCMNG
                     AALCALQHFMPELVERANPEVGTEDWEHLEILQKEDVKPVLVNGVNEAVEESEFTLSEYTT
                     SYGIEHAQCVPPSEFSFPSFITESYQTTHPTSSPFTLSEKYENDYPSVTLRDPLQTDT
                     LQNDYFATKQRVVTPDNMCMGRTSPGSGPTQLWQFLLELLTDKSCQSFISWTGDEWEF
                     KLSPPDFVARFWGKPFENFFEMNYFELSPCLRYYYIPENIIHECACERYVYREVCDLQSL
                     LGYTPEELHAMLDVKPDADE"
     intron          1009. .1269
                     /note="alternatively spliced exon"
BASE COUNT      418 a    443 c    408 g    335 t
ORIGIN


alignment_scores:
            Quality:  251.50              Length:     540
              Ratio:    1.251               Gaps:      22
  Percent Similarity:   47.925   Percent Identity: 21.698

alignment_block:
 US-09-126-9458-2 x HSCETS1   ..

 Align seq 1/1  to: HSCETS1  from: 1  to: 1604

       24 SerArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluAr 40
          :::|||::: ||   |||:::   |||   |||:::
       10 GCGAGGGCCGGGCAGGAGGACCGGGCGCGCGGCGGG.............. 45

       40 qArgAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerA 57
          |||   |||:::|||||        ::::::|||||||||
       46 .CGAGGCTGGGACCGAGCGCGCTCACTTCGCCGCAAAGTGCC..... 87

       57 laPheTyrLeuSerTyrPheAspMetLeuTyrProGluAspSerSerTrp 73
                                                  :::   |||
       88 ..................................AACTTCCCCTGG 99

       74 AlaAla...LysAlaProGly..AlaSerSerArgGluGluPro..... 86
          ::|||   :::|||||||::  |||   ::::::|||:::
      100 AGTGCCGGGCGCGCACCGTCCGGGCGCGGGGGAAAGAAAGGCAGCGGAA 149

       87 ..............................ProGluGluProGluGlnCy 93
                                        |||   :::|||::   ||
      150 TTGGAGATTTTGGGAAGAAAGTCGGATTCCCCCGTCCCCTCCCCCTG 199

       93 s.............................................P 94
          |                                            |
      200 TTACTAATCCTCATTAAAAAGAAAAACAACAATAACTGCAAACTTGCTAC 249

       94 roValIleAspSerGlnAlaProAla.....GlySerLeuAspLeu 107
          ||          |||   :::|||:::         :::::::|||||
      250 CATCCCGTACGTCCCCCACTCCTGCCACCATGAAGGCGGGCCGTCGATCTC 299

      108 ValPro..................................... 109
          |||
      300 AAGCCGACTCTCACCATCATCAAGACGGAAAAAGTCGATCTGGAGCTTTT 349

      110 .....................GlyGlyLeuThrLeuGluGluHisSerL 119
                               :::  :::   |||         |||
      350 CCCCTCCCCGGATATGGAATGTGCAGATGTCCCACTATTAACTCCAAGCA 399

      119 euGluGluValGlnSerMetValValGlyGluValLeuLysAspIleGlu 135
          ::::::::::|||:::::::::    :::::::::   :::
      400 GCAAACAAATCATCTCTCAACGCATTAAAAGCTACTTTCAGTGGTTTCACT 449

      136 ThrAlaCysLysLeuGlnAspIleThrAlaAspProMetAspTrpSerPr 152
                :::   |||:::|||      ||||||   :::|||:::
      450 AAACAACAGCAACGACTGGGGATCCCAAAAGACGCCCCGGCAGTGGACAGA 499

      153 oSerAsnValGlnLysTrpLeuLeuTrpThrGluHisGlnIyrArgLeuP 169
          :::::|||:::   |||:::::|||:::   :::::::   |||
      500 AACCCATGTTCGGGACTGGGTGATGTGGGCTGTGAATGAATTCAGCCTG. 548

      169 roProMetGlyLysAlaPheGlnGlu.....LeuAlaGlyLysGluLeu 185
                 |||      |||||:::      :::   |||     |||
      549 .    AAAGGTGTAGACTTCCAGAACTTCTGTATGAATGGGAGCAGCCCTG 594

      186 CysAlaMetSerGlyGluGlnPheArgGlnArgSerPro   LeuGlyGl 199
          ||||||:::::::::::::::   |||   :::   :::|||   :::   ||
      594 TGCGCCCTGGGTAAAGACTGCTTTCTCGGAGCTGGCCCCAGACTTTGTTGG 643

      199 yAspValIleHisAlaHisLeuAspIleTrpLys.................. 210
          |||:::|||      |||||:::|||   :::
      644 GGACATCTTATGGCAACATCTACAGATCCTGCACAAAGAGGATGTGAAAC 693

      211 .....................SerAlaAlaTrpMetLysGluArg...Thr 219
                              :::   |||:::   :::::|||   |||
      694 CATATCAACGTTAATGGAGTCAACCCAGCTATCCAGAATCCCGCTATACC 743

      220 SerProGlyAlaIleHisTyr.................CysAlaSerPh 240
          |||      |||   |||                   ||||::
      744 TCGGATTACTTCATTAGCTATGGTATTGAGCATGCCCAGTGTGTTCCACC 793

      240 rSerGlu.......................GluSerTrp.......... 245
          ||||||                      |||||||:::
      794 ATCGGAGTTCTCAGAGCCCAGCTTCATCACAGAGTCCTATCAGACGCTCC 843

      245 ..................................................... 245

      844 ATCCCATCAGCTCGGAAGAGCTCCTCTCCCTCAAGTATGAGAATGACTAC 893
```

```
      240 ..............................ThrAspSer............ 248
                                       |||||||:::
      894 CCCTCGGTCATTCTCCGGAGACCCTCTCCAGACAGACACCTTGCAGAATGA 943

      249 ...................GluVal.........AspSerSerCysSerG 266
                            ||||||         |||:::   |||   |
      944 CTACTTTGCTATCAAACAAGAAGTCGTCACCCCAGACAACATGTGCATGG 993

      246 lysGln.................................................. 247
          ||:::
      994 GGAGGACCAGTCGTGGTAAACTCGGGGGCCAGGACTCTTTTGAAAGCATA 1043

      247 ..................................................... 247

     1044 GAGAGCTACGATAGTTCGATCGCCTCACCCAGTCCTGGAGCAGCCAGTC 1093

      247 ..................................................... 247

     1094 ATCTTTCAACAGCCTGCAGCGTGTTCCCTCCTATGACAGCTTCGACTCAG 1143

      247 ..................................................... 247

     1144 AGGACTATCCGGCTGCCCTGCCCAACCACAAGCCCAAGGGCACCTTCAAG 1193

      247 ..................................................... 247

     1194 GACTATGTGCGGCACCGTGCTGACCTCAATAAGGACAAGCCTGTCATTCC 1243

      248 ..............................ProIleHisLeuTrpC 258
                                       ||||||:::|||||||
     1244 TGCTGCTGCCCTAGCTGCCTACACAGGCAGTGGACCAATCCAGCTATGGC 1293

      259 lnPheLeuLysGluLeuLeuLeuLysProHisSerTyrGlyArgPheIle 269
          |||||||   |||   ||||||          |||         ||||||
     1294 AGTTTCTTCTGGAA...TTACTCACTGATAAATCCTGTCAGTCTTTTATC 1340

      270 ArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAspSerAlaGl 289
          |||   :::   :::          ||||||:::::|||       ::
     1341 AGCTGGACAGGA...GATGGCTGGGAATTCAAACTTTCTGACCCAGATGA 1387

      286 nValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMetAsnTyrA 303
          :||||||||||   ||||||   ||||||||||:::|||   ||||||||||
     1388 GCTGGCCAGGAGATGGGGAAAGAGGAAAAACAAACCTAAGATGAATTATG 1437

      303 spLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyIleIleArg 319
          ::||||||||||||:::::|||   ||||||   |||:::||||||::
     1438 AGAAACTGAGCCGTGGCCTACGCTACTATTACGACAAAAACATCATCCAC 1487

      320 LysProAspIleSerGluArgLeuValTyrGlnPheVal 332
          |||         ::::::|||   ||||||:::||||||
     1488 AAG...ACAGCGGGGAAACGCTACGTGTACCGCTTTGTG 1523

seq name: qb pr%:AF116508

seq documentation block:
LOCUS       AF116508      1592 bp    mRNA             PRI       19-MAR-2001
DEFINITION  Homo sapiens Ets transcription factor TEL 2a (TEL 2) mRNA, complete
            cds.
ACCESSION   AF116508
VERSION     AF116508.1  GI:5442095
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1592)
  AUTHORS   Gu,X., Shin,B.H., Akbarali,Y., Weiss,A., Boltax,J., Oettgen,P. and
            Libermann,T.A.
  TITLE     Tel 2 is a novel transcriptional repressor related to the ets
            factor tel/etv 6
  JOURNAL   J. Biol. Chem. 276 (12), 9421 9436 (2001)
  MEDLINE   21154768
REFERENCE   2  (bases 1 to 1592)
  AUTHORS   Shin,B.H., Boltax,J., Gu,X., Akbarali,Y. and Libermann,T.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC 1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1. .1592
                     /organism "Homo sapiens"
                     /db_xref "taxon:9606"
                     /chromosome "6"
                     /map "6p21"
                     /tissue_type "pancreas"
     gene            1. .1592
                     /gene "TEL 2"
     misc_feature    76. .411
                     /gene "TEL 2"
                     /note "Ets transcription factor TEL 2a pointed domain;
                     possibly translated; peptide identified by in vitro
                     translation assay"
     CDS             294. .1142
                     /gene "TEL 2"
                     /note "DNA binding domain"
                     /codon_start 1
                     /product "Ets transcription factor TEL 2a"
                     /protein_id "AAD43250.1"
                     /db_xref "GI:5442096"
                     /translation "MRRGAKVRQERIRQLHRHQQGKLPAPCAQLRSGERKIGEPLSGNPC
                     DVLYELLQYIKIQRKALVCGEEEGGTEEQLKTIPQHSPVPPEEVTQPSQMDTRRQHLQ
                     PPDPGLTSNFGHLDDPGLARWTPGKEESLNLCBCAELGCRTQGVCSFPAMPQAPTHGR
                     IADCRLLWDYVVQLLLDTKYEPYIKWEDKSAKIFKVVDPNCLARLWCNBKNKVNMIYE
                     KMSKALPHIKELNKEEQKKLIEKELEELPSEMVQEHSSRLEETLESQQQQYLKPPE
                     KPEISP"
BASE COUNT       381 a    467 c    466 g    278 t
ORIGIN


alignment scores:
            Quality:  251.00            Length:     351
              Ratio:  1.549               Gaps:      12
 Percent Similarity:  46.154   Percent Identity: 24.786

alignment block:
 US-09-126-945B-2 x AF116508   ..

 Align seq 1/1  to: AF116508  from: 1  to: 1592

      76 LysAlaProGlyAlaSerSerArgGluGluProProGluGluProGluGl 92
         :::   |||   ||||||||          ||||||          |||
      43 AGAAAGCCGGTGGCTTCCTCA.....CCTCACCTGTAATGCAGGAGGA 86

      92 nCysProValIleAspSerGlnAlaProAlaGlySerLeuAspLeuValP 109
            ::||||::      :::|||:::::::::::::      |
      87 AGAATTGGCTATTTCTCCTATAAGCCCTGTGGCAGCCATGCCT.....C 130

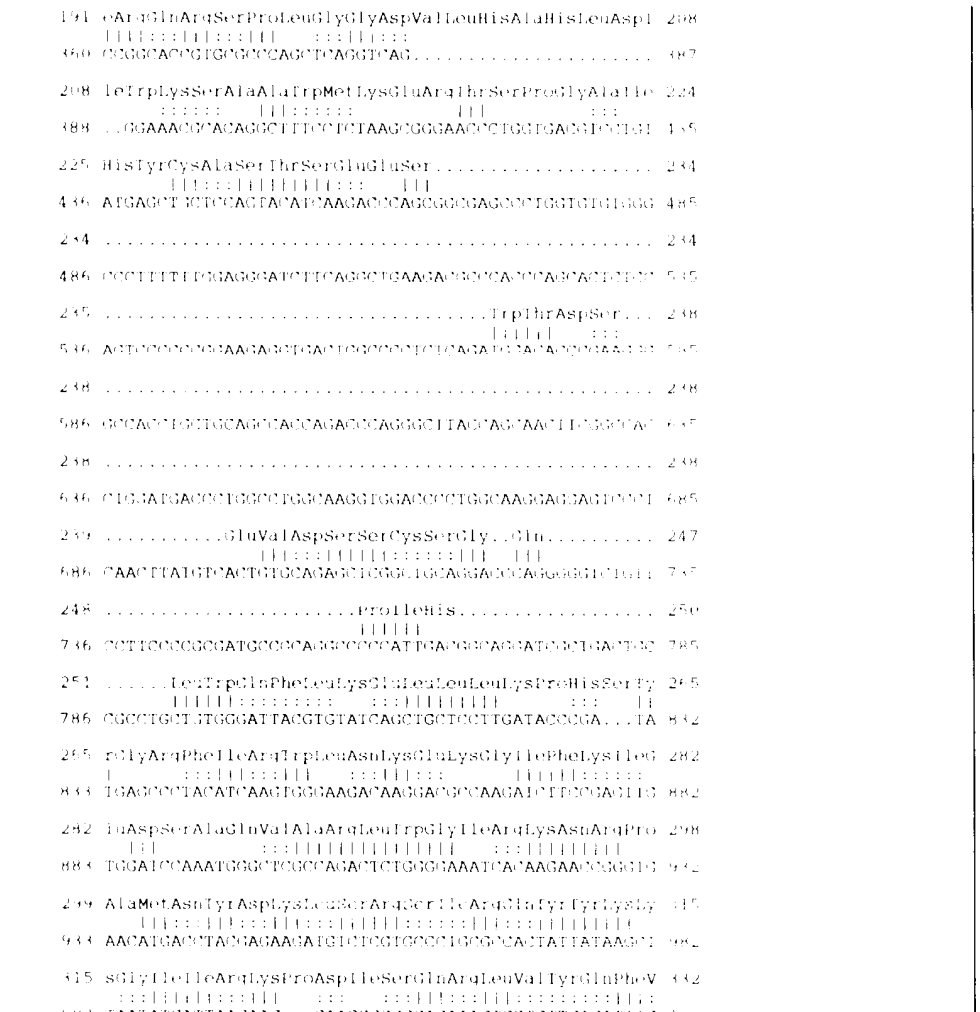
     109 roGlyGlyLeuThrLeuGluGluHisSerLeuGluGlnValGluSerMet 125
         ||   |||      ::::::   :::         |||::::::   :::
     131 CCCTAGGCACCCACGTGCAAGCCAGATGTGAAGCTCAAATTAACCTGCTG 180

     126 ValValGlyGlnValLeuLysAspIleGluThrAlaCysLysLeuLeu.. 141
             |||   :::                        |||||||||
     181 GGTGAAGGGGGAATC.....................TGAAGCTGCCAGG 209

     142 .AsnIleThrAlaAspProMetAspTrpSerProSerAsnValGluLysL 158
          ::::::      :::|||      ||||||   :::::|||      |
     210 AAGACTCCCCATCCAGCCCCCACTCTGGAGCAGGGAGGACGTGCTGCACT 259

     158 rpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMetGlyLysAla 174
         |||||   ||:::|||:::::::|||   ||||||      ::::::
     260 GGCTGCGCTGGGCAGAGCAGGAGTACCTCTCTGCCATGCACCGCGGAGCAC 309

     175 PheGlnGluLeuAlaGlyLysGluLeuCysAlaMetSerGlnGluGlnPh 191
             ||:::   |||:::   ||||||   :::::::::::::::||
     310 GGGTTCGAGATGAACGGACGCGCCTCTGCATCCTCACCAAGGACGACTT 359
```



```
seq_name: qb_rol:MMETS1

seq_documentation_block:
LOCUS       MMETS1       1902 bp    mRNA            ROD       13-MAY-1992
DEFINITION  M.musculus ets-1 mRNA.
ACCESSION   X53953
VERSION     X53953.1  GI:50870
KEYWORDS    ets-1 gene; proto-oncogene; transcription factor.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1902)
  AUTHORS   Graves,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-1990) Graves B., University of Utah, Dept. of
            Cellular, Viral & Molecular Biology, Univ. of Utah Medical Center,
            MC5C384, Salt Lake City, Utah 84132, USA
REFERENCE   2  (bases 1 to 1902)
  AUTHORS   Gunther,C.V., Nye,J.A., Bryner,R.S. and Graves,B.J.
  TITLE     Sequence specific DNA binding of the proto oncoprotein ets-1
            defines a transcriptional activator sequence within the long
            terminal repeat of the Moloney murine sarcoma virus
  JOURNAL   Genes Dev. 4 (4), 667-679 (1990)
  MEDLINE   90299137
COMMENT     Homolog of human and chicken ets-1 gene. Related to E26 avian
            transforming virus oncogene.
FEATURES             Location/Qualifiers
     source          1. .1902
                     /organism="Mus musculus"
                     /strain "balb/c"
                     /db_xref "taxon:10090"
                     /dev_stage "2 week old"
                     /tissue_type "thymus"
     gene            510. .1832
                     /gene "ets-1"
     CDS             510. .1832
                     /gene "ets-1"
                     /codon_start 1
                     /product "proto oncogene"
                     /protein_id "CAA37904.1"
                     /db_xref="GI:50871"
                     /db_xref "SWISS-PROT:P27577"
                     /translation "MRAAVDLKPSLSTFKDYVRPSLEPEPSQFIMEVAEYFKQLEAE
                     MSQALKAIESGFTKEQQRKLGTPKDPPKQWTETHVRDWVMWAVNEPSLKGVDFQKFCMSG
                     AALAVALQKEGTELPLAPDFVGRTLWEHLEDLGQKEDVKFYQVNGANPIYPESCYISDYFI
                     SYGIEHAQCVPPSEFSEPSFITESYQTLHPISSFEELISLKYENDYPSVIIQLEPLGIDI
                     TQTDYFATKQFVLTPDNMCLGRASKCKIGGQDSFFSVFSYDSCDRLTQSWSSQSSFNS
                     LQRVPSYDSFEDYEDYPAALPNHEPSACIPEAYYRKQPAILNKNQKEVIPAAALAGYIGSP
                     IQIWQFLLELLTDKSCQSFISWTGDGWEFKLSDPDEVARRWGKRKNKPKMNYEKLSRG
                     LRYYYDKNIIHKTAGKRFVYREVCDGKSLLGYIPEFIHAMLQVKPDAD"
BASE COUNT      496 a    500 c    453 g    453 t
ORIGIN


alignment_scores:
            Quality:  250.50              Length:    499
              Ratio:  1.354                 Gaps:     22
 Percent Similarity:  87.074   Percent Identity: 23.447

alignment_block:
 US-09-126-945B-2 x MMETS1   ..

 Align seq 1/1  to: MMETS1  from: 1  to: 1902

       8 LeuSerSerValSerProSerHisLeuLeuProProAspIleValSe 24
         :::   |||:::       |||   |||   |||   |||   :::   :|
     447 AlACACAGIAIAGIGAGAAGCAAACTIACICICAICCCGCCGGCCCCCIC 496

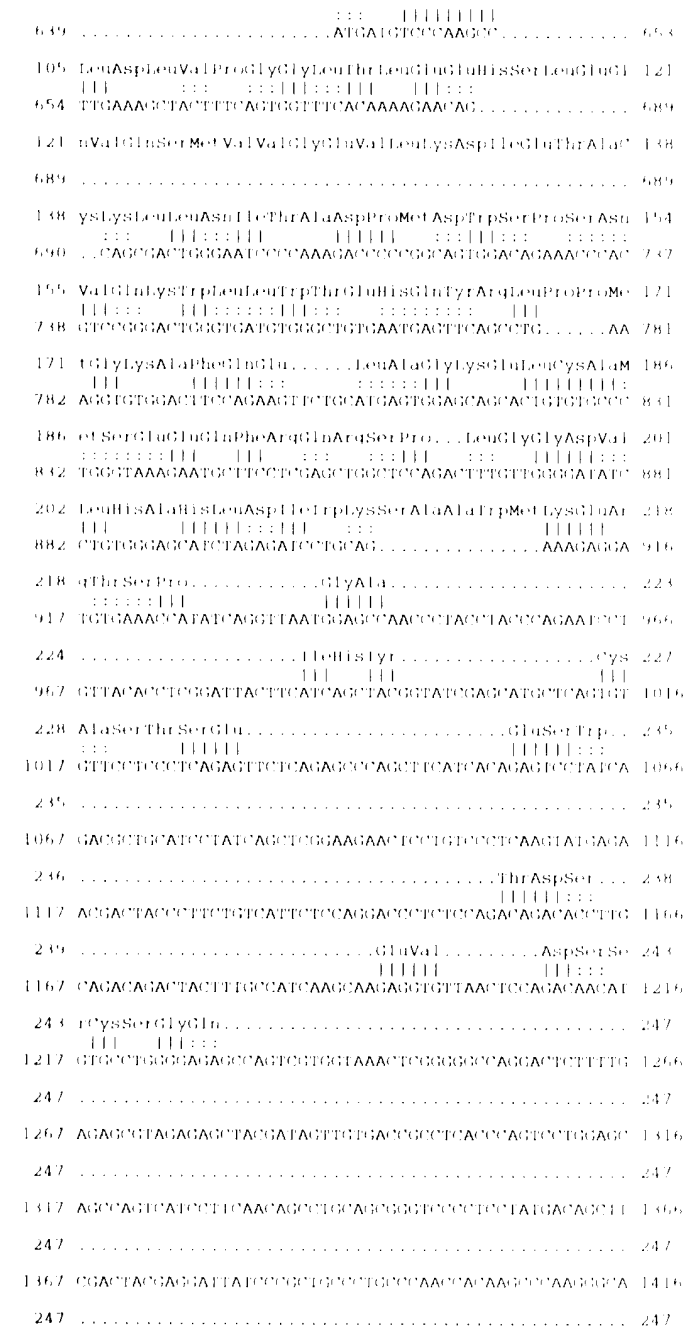
      24 rArgThrGlyLeuGluLysAlaAlaAlaGlyAlaValGlyLeuGluArgA 41
         |      |||      ||||||||||:::
     497 AACICCGGGCACCAIGAAGCCGGCCGIC..................... 524

      41 rgAspTrpSerProSerProProAlaThrProGluGlnGlyLeuSerAla 57
           |||   :::|||:::                          ||||:::
     525 ..GAICICAAGCCGACI.....................CICACCAIC 548

      58 PheTyrLeuSerTyrPheAspMet...LeuTyrProGluAspSerSerIl 74
         :::       :::       |||:::   |||:::|||:::   :::
     549 AICAAGACAGAAAAAGIGGAICICGAGCIIICCCIICCCCGGACAIGGA 598

      73 pAlaAlaTysAlaPro.....GlyAlaSerSerArgGluGluProProG 88
         :::|||   :::|||           |||||:::|||
     599 APGTGCAGARGICCCGCIGIIAACICCGAGCAGCAAAGAA.......... 638

      88 lnGlnProGluGlnCysProValIleAspSerGlnAlaProAlaGlySer 104
```


649 ATGATGTCCCAAGCC 663
105 LeuAspLeuValProGlyGlyLeuThrLeuGluGluHisSerLeuGluGl 121
654 TTGAAAGCTACTTTCAGTGGTTTCACAAAAGAACAG 689
121 nValGlnSerMetValValGlyGluValLeuLysAspIleGluThrAlaC 138
689 689
138 ysLysLeuLeuAsnIleThrAlaAspProMetAspTrpSerProSerAsn 154
690 CAGCCACTGGGAATCCCCAAAGACCCCCGGCAGTGGACAGAAACCCAC 737
155 ValGlnLysTrpLeuLeuTrpThrGluHisGlnTyrArgLeuProProMe 171
738 GTCCGGGACTGGTGATGTGGGCTGTGAATGAGTTCAGCCTG.....AA 781
171 tGlyLysAlaPheGlnGlu.....LeuAlaGlyLysGluLeuCysAlaM 186
782 AGGTGTGGACTTCCAGAAGTTCTGCATGAGTGGAGCAGCACTGTGTGCCC 831
186 etSerGluGluGlnPheArgGlnArgSerPro...LeuGlyGlyAspVal 201
832 TGGGTAAAGAATGCTTCCTCGAGCTGGCTCCAGACTTTGTTGGGGATATC 881
202 LeuHisAlaHisLeuAspIleTrpLysSerAlaAlaTrpMetLysGlnAr 218
882 CTGTGGGAGCATCTAGAGATCCTGCAG.....AAAGAGGA 916
218 qThrSerPro.....GlyAla..... 223
917 TGTGAAACCATATCAGGTTAATGGAGCCAACCCTACCTACCCAGAATCCT 966
224 .....IleHisTyr.....Cys 227
967 GTTACACCTCGGATTACTTCATCAGCTACGGTATCGAGCATGCTCAGTGT 1016
228 AlaSerThrSerGlu.....GluSerTrp.. 235
1017 GTTCCTCCCTCAGAGTTCTCAGAGCCCAGCTTCATCACAGAGTCCTATCA 1066
235 235
1067 GACGCTGCATCCTATCAGCTCGGAAGAACTCCTGTCCCTCAAGTATGAGA 1116
236 .....ThrAspSer... 238
1117 ACGACTACCCTTCTGTCATTCTCCAGGACCCTCTCCAGACAGACACCTTC 1166
239 .....GluVal.....AspSerSe 243
1167 CAGACAGACTACTTTGCCATCAAGCAAGAGGTGTTAACTCCAGACAACAT 1216
243 rCysSerGlyGln..... 247
1217 GTGCCTGGGGAGAGCCAGTCGTGGTAAACTCGGGGCCCAGGACTCTTTTG 1266
247 247
1267 AGAGCGTAGACAGCTACGATACTTGTGACCGCCTCACCCAGTCCTGGAGC 1316
247 247
1317 AGCCAGTCATCCTTCAACAGCCTGCAGCGGGTCCCCTCCTATGACAGCTT 1366
247 247
1367 CGACTACGAGGATTATCCCGCTGCCCTGCCCAACCACAAGCCCAAGGGCA 1416
247 247
1417 CCTTCAAGGACTATGTGCCTGACCGTGCTCACCTCAACAAGGACAAGCCT 1466
248 .....ProIleHi 250
1467 GTCATCCTGCTGCTGCCCTGGCTGGCTACACAGGAAGTGGGCCGATCCA 1516
250 sLeuTrpGlnPheLeuLysGluLeuLeuLysProHisSerTyrGlyA 267
1517 GCTGTGGCAGTTTCTTCTGGAA...TTACTCACTGATAAGTCTTGTCAGT 1564
267 rqPheIleArgTrpLeuAsnLysGluLysGlyIlePheLysIleGluAsp 283
1564 CCTTTATCAGCTGGACAGGA...GATGGCTGGGAATTCAAGCTTTCTGAC 1610
284 SerAlaGlnValAlaArgLeuTrpGlyIleArgLysAsnArgProAlaMe 300
1611 CCAGATGAGGTGGCCAGGAGATGGGGAAAGAGGAAAAACAAACCTAAGAT 1660
300 tAsnTyrAspLysLeuSerArgSerIleArgGlnTyrTyrLysLysGlyI 317
1661 GAATTATGAGAAACTGAGCCGTGGCCTTCGCTACTATTATGACAAAAATA 1710
317 leIleArgLysProAspIleSerGlnArgLeuValTyrGlnPheVal 332
1711 TCATCCACAAG...ACGGCGGGCAAGCGCTACGTATACCGCTTTGTG 1754